# A Machine Learning-Based Non-Contact Respiratory Rate Monitoring Method Using an RGB Camera

Mohammad Ghodratigohar

Thesis submitted in partial fulfillment of the requirements for the
**MASTER OF APPLIED SCIENCE**
IN BIOMEDICAL ENGINEERING

Ottawa-Carleton Institute for Biomedical Engineering School of
Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada

# Abstract

Non-contact measurement of Respiratory Rate (RR) is suitable for various clinical and home-based healthcare applications. RR monitoring can inform healthcare providers of early indicators of critical illnesses. However, the obtrusive nature of contact-based sensors for RR monitoring makes them uncomfortable for extended use and vulnerable to movement-derived noise. Hence, camera-based approaches have attracted considerable attention as they enable contact-free RR monitoring. This thesis presents an improved non-contact method for RR monitoring that leverages camera derived remote photoplethysmography (rPPG) to measure RR. Unlike previous work, the proposed method supports subject movement during monitoring.

We apply Independent Component Analysis (ICA) on the RGB channels of facial videos to distinguish the source (i.e. PPG signal) from noise. We use the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) scheme to decompose the selected ICA output into its Intrinsic Mode Functions (IMFs). We propose a Machine Learning (ML) algorithm to select the IMF that best reflects the RR. We evaluated the proposed method on 200 facial videos collected from 10 subjects. Our approach decreased the RMSE by at least 39.6% compared to state-of-the-art techniques when subjects were stationary. For subjects in movement, we achieved an RMSE of 2.30 BPM (breaths/min).

# Acknowledgments

The words are not capable of describing my greatest appreciation to my supervisor Dr. Hussein Al Osman. I have been extremely lucky to have a supervisor who cared so much about my work. During the last two years, I benefited from his guidance, great support, fruitful discussions and kind pieces of advice that encouraged me to do my best. It was a real privilege and an honor for me to be supervised with his exceptional scientific knowledge and his extraordinary human qualities. He has always been responsive to my questions with his patience and enthusiastically and guided me toward the best solutions. He taught me how to do and think as an academic researcher. He has kindly provided me with detailed feedback on my research paper and on my thesis by dedicating his valuable times. It would never have been possible for me to take this work to completion without his incredible support, encouragement, and trust.

I am truly grateful to my parents for their immeasurable love and care. They have always encouraged me to believe in my potential and pursue my dreams. They are absolutely the root of my life and they helped me a lot to reach this stage in my life and I cannot really express how important their presence has been throughout my life. Also, I would like to thank my only sister for everlasting encouragement and support.

I would like to thank all of my friends and lab colleagues for their participation and help in my research. I highly appreciate their cooperation and letting me have these wonderful friends.

## Dedication

To My Parents

And Sister

# Table of Contents

# Table of Figures

# Table of Tables



# Glossary of Terms

**RR:** Respiratory Rate

**HR**: Heart Rate

**HRV**: Heart Rate Variability

**ICA**: Independent Component Analysis

**ML**: Machine Learning

**CEEMDAN :** Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

**EMD :** Empirical Mode Decomposition

**EEMD :** Ensemble Empirical Mode Decomposition

**IMF :** Intrinsic Mode Function

**ECG**: Electrocardiography

**PPG**: Photo-plethysmography

**rPPG**: Remote Photo-plethysmography

**EVM**: Eulerian video magnification

**BVP**: Blood volume pulse

**SpO2:** Blood oxygen saturation

**RMSE**: Root Mean Square of Error

**LF**: Low frequency

**HF**: High frequency

**PSD**: Power spectral density

**FFT**: Fast Fourier Transform

**FIR**: Finite impulse response

**IIR**: Infinite impulse response

**RGB**: Red, Green, Blue

**SNR**: Signal to Noise Ratio

**HSL**: Hue, Saturation, Lightness

**ROI**: Region of Interest

**JADE**: Joint Approximate Diagonalization of Eigen matrices

**FPS**: Frame per second

**CNN**: Convolutional Neural Network

**ANN**: Artificial Neural Network

# Chapter 1. Introduction

Respiratory Rate (RR), commonly defined as the number of breaths per minute, is an important indicator of the health status and aids in identifying health abnormalities [1]. The RR measure can serve as a predictor of critical illnesses. For instance, RR is screened in emergency departments of hospitals for indicating respiratory dysfunctions and predicting cardiac arrest [1]. In primary care, RR can identify apnea, pneumonia and pulmonary embolism [1]. Hence, monitoring RR can support effective triage decisions and facilitate the regionalization of critical care [2]. Its continuous measurement is prominent for monitoring health conditions and it is typically measured as the number of breaths per minute (BPM).

## 1.1. Problem Statement

The conventional techniques to measure respiration rate are contract-based and require sensors to be attached to the body, e.g. belt, or facemask [3]. These sensors limit the movement of the patient. Even conventional ambulatory sensors designed for home care affect the freedom of movement due to excessive wiring. Moreover, in long-term monitoring, the sensor attachment to the body may lead to skin irritations and/or patient discomfort. For example, monitoring vital signs of burned skin patients or premature infants in the Neonatal Intensive Care Unit (NICU) requires the use of adhesives to attach sensors to their skin which might result in pain and skin irritation [5]. Although some contactless methods, namely air mattress sensors, are capable of monitoring respiration [4], the RR signal can be distorted by non-respiratory-induced movements [4]. Furthermore, these sensors are restricted to particular contexts (e.g. subject sleeping on the mattress). Hence, non-contact techniques are well suited for these scenarios.

## 1.2. Goals and Objectives

Lately, researchers have been focusing on the feasibility of measuring RR using camera through motion vectors related to the respiratory-induced movements and remote Photoplethysmography (rPPG). Such remote measurement solutions can be applicable to situations where contact-based sensors are not desirable or feasible such as the monitoring of infants in neonatal intensive care, elderly individuals in senior care centers, patients in hospital emergency waiting rooms, and prisoners on suicide watch.

The measurement of rPPG is based on the principle that when the heart pumps blood, the arterioles blood volume increases which leads the skin color to vary. Therefore, a standard RGB camera can record these color changes instigated by the light absorption properties of the blood. Consequently, we have developed a method to extract RR from a facial video of a subject in movement. The objective of this study is to assess the performance of the aforementioned method by comparing the RR extracted from a facial video with those collected via capacitive belt sensor and self-counting as ground truth.

Also, we note that previous video-based RR measurement studies present several shortcomings. Firstly, individuals were limited to slight movements [6][7][8]. Specifically, for motion-based measurement techniques [9-10], it is challenging to differentiate between respiratory-induced motions and other movements that are unrelated to breathing. Moreover, the measured rPPG signal can be corrupted by movement artifacts and noise derived from lighting changes [11]. The RR is subtly modulated in the PPG signal. Hence, its accurate extraction is highly affected by noise [12]. Therefore, we focus on improving the robustness of rPPG-based RR measurement while relieving

the assumption of subject stillness. Hence, the subject can freely move within the environment. The proposed algorithm can estimate the RR if it receives a facial video containing the subject's forehead (given that both eyes are visible).

## 1.3. Motivation

Contactless techniques to monitor vital signs have a variety of applications in clinical areas and home-based healthcare. The common techniques to measure RR are contact-based which requires sensors attached to the body [3] while in long-term monitoring, these attachments may lead to discomfort and limitations in mobility. Furthermore, in some cases, patients refuse continuous monitoring and hence reject obtrusive sensors (e.g. prisoners on suicide watch).

We can envision numerous scenarios where remote measurement of RR can be desirable and effective:

- Monitoring residents of senior care centers to alert staff to emergencies
- Monitoring patients in hospital emergency waiting rooms to notify staff if a patient's condition unexpectedly deteriorates
- Monitoring prisoners on suicide watch to communicate emergencies to guards and medical staff
- Monitoring sleeping infants to send an alarm to parents in case of an emergency

We present a novel motion-robust non-contact camera-based method as a replacement to contact-based conventional sensors for extraction of respiratory rate. Typical cameras are surrounding us everywhere. From security cameras in public places to personal ones on and off mobile phones. Therefore, the ability to measure this vital sign using a typical and widely available camera, with no additional hardware, can prove favorable for the diagnosis of a number of important clinical conditions.

## 1.4. Contributions

We summarize our contributions as follows:

1. We apply the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) [13] to decompose the rPPG signal and obtain the most informative coefficient that reflects RR.
2. We develop a Machine Learning (ML) algorithm to select the most informative CEEMDAN derived IMF.

### 1.4.1 Publication

The outcome of this thesis was submitted as a journal paper to the IEEE Journal of Translational Engineering in Health and Medicine:

M.Ghodratigohar, H.Ghanadian, H. Al Osman. "A Non-Contact Respiratory Rate monitory method using CEEMDAN and Machine Learning". Submitted to Journal of Translational Engineering in Health and Medicine (JTEHM), Submitted March 30th, 2019.

## 1.5. Thesis Outline

The rest of this thesis is organized as follows:

**Chapter 2** discusses the relevant background and related work. We summarize previous literature related to the contactless approaches to monitor the RR signal.

**Chapter 3** describes the proposed method to estimate the RR signal with a camera. Furthermore, we describe the proposed decomposition methods and machine learning schemes.

**Chapter 4** presents the results for the validation and lab experiments we conducted to assess the effectiveness of the proposed method. Moreover, we discuss the results and present our conclusions.

**Chapter 5** summarizes the thesis findings and provides insights into future wor

# Chapter 2. Background and Related Works

In this chapter, we provide a brief explanation of the relevant concepts in machine learning, image processing, and physiological signals. Moreover, we present existing non-contact RR measurement methodologies and summarize the rPPG based methods in Table 1 and Table 2.

## 2.1. RR Measurement

RR, defined as the number of respirations per unit of time, is a significant vital sign measure for assessing individuals' well-being [1]. Its regular measurement is crucial in supporting the triage decision and facilitate regionalization of critical care [2]. Fluctuations in RR can be an indicator of potentially serious clinical event such cardiac arrest and the basis for admission to the intensive care unit [14-17]. The process of measuring RR can be classified into either contact or non-contact method.

### 2.1.1 Contact Measurement

The common techniques to measure respiration rate are contact-based which requires sensors attached to the body, e.g. belt or a facemask [3]. They are usually based on measuring one of the following parameters: respiratory related sounds, airflow, chest or abdominal movements, respiratory CO2 emission, pulse oximetry (SpO2-blood oxygen Saturn) and Electrocardiogram (ECG) [3].

#### 2.1.1.1 Acoustic Monitoring

In the acoustic based methods, a microphone placed close to the airways records respiratory sounds and through its variations the RR can be estimated using frequency analysis [18]. In [19], they extracted

breaths sounds from an acoustic signal contaminated by environmental noise using spectral characteristics of normal breaths sounds measured at a distance of 100 mm. They observed that environmental noise has unique frequency distribution patterns. Hence, they can distinguish breaths sounds by filtering for the appropriate range. Werthammer et al. [20] reported that snorting, speaking, crying, coughing, etc. had a negative effect on the function of this method.

#### 2.1.1.2 Airflow Monitoring

Since exhaled air is warm and humid, RR can be measured by monitoring the respiratory airflow [21]. A nasal or oral thermistor can be used to detect the fluctuations of temperature between inhaled and exhaled air to measure the airflow and RR. This method is vulnerable to the high incidence of thermistor displacement [22]. Some patients may not feel comfortable with the sensor especially that it is placed over the mouth and/or upper lip area [23].

#### 2.1.1.3 CO2 Monitoring

The CO2 monitoring method measures an overall estimate of changes in CO2 levels by diffusing gas with a heated electrode set to the skin. However, the electrode might cause burns on the sensitive neonatal skin [24]. Folke et al. [25] also propose a CO2 sensor that measures RR. However, they indicated that subtle design changes in the collecting device could introduce large differences in the sensor performance [23].

#### 2.1.1.4 Movement Detection

For chest and abdominal derived movement detection, a sensor integrated into a belt wrapped around the chest or abdomen is used. A conductor can be integrated into the belt such that during respiration, when the chest area expands, the conductor resistance increases. The variations in the conductor's resistance reflects RR. A strain gauge can be used with a belt device wrapped around the patient's chest to measure changes in the thoracic or abdominal circumference during breathing which

corresponds to respiration cycles [25]. In inductive plethysmography, a sinusoidal wire is wrapped around the chest and/or abdomen. The sensor measures variations in inductance related to changes in the shape of the torso caused by respiration [26]. A piezoelectric sensor integrated within a belt wrapped around the chest generates a small voltage when it stretches due to respiration derived changes in the chest or abdomen volume. Hence, the RR information can be estimated from generated voltage [26]. Kang et al. [27] developed a capacitive based respiration sensor integrated with a belt. The sensor measures the alteration in the area between two capacitor plates that slide parallel to one another. During respiration, the sensor will be stretched and since the plates are laterally connected with one-dimensionally stretchable nonwovens, it enables the plates to slide in opposite directions. Therefore, the RR can be obtained based on respiration derived moment of the plates inside the capacitive sensor.

#### 2.1.1.5 PPG

PPG-based measurement is another low-cost and non-invasive means of sensing the blood volume pulse (BVP) that leads to estimating RR. The PPG signal is obtained from variations in the reflected light from the tissue [28] through a pulse oximeter attached to the finger to measure the oxygen saturation level in the blood. During exhalation, the parasympathetic nervous system increases the flexibility of blood vessels which affects the pulse wave of the PPG signal [29]. Respiration modulates the amplitude and rate of PPG pulses. Therefore, the respiratory information can be extracted from the pulse width, rate, and amplitude variability of the PPG signal [29]. Leonard et al. [30] processed the PPG signals of ten healthy adults using a wavelet transform that led to the observation of respiration waveforms. Wertheim et al. [31] demonstrated the feasibility of extracting RR from the PPG of infants by detecting the peak of the PPG frequency spectrum associated with RR. Lazaro et al. [32] recorded a PPG signal using a smartphone (iPhone 4s). The subjects were asked to place their finger on the camera as the onboard flashlight illuminated the

subjects' fingers. They extracted the PPG signals by averaging the green channel on a 50x50 pixel Region of Interest (ROI) selected from the captured video. They then combined derived respiration information from the pulse, rate and amplitude variability of the PPG signal to estimate the RR.

#### 2.1.1.6 ECG

ECG derived RR is based on the fact that respiration modulates the ECG signal [3]. The body-surface ECG is affected by the electrodes' motions associated with respiration. Such motions lead to fluctuations in the thoracic electrical impedance during inhalation and exhalation. Therefore, fluctuations in the mean cardiac electrical axis are correlated with respirations which permits the estimation of RR [33]. Mazzanti et al. [34] utilized Principal Component Analysis (PCA) to estimate the ECG leads that contributed the most to the respiratory variations. The shortcoming of these methods is that movement artifacts cause errors in RR estimation [3].

### 2.1.2. Non-Contact Measurement

#### 2.1.2.1 Radar

To the best of our knowledge, Caro et al. [35] were the first to introduce remote respiration monitoring using radar as an apnea (i.e. abrupt stop in respiration) detector. They investigated the existence of a Doppler shift in the frequency of the received wave to measure breathing. Furthermore, Chen et al. [36] utilized a Doppler radar [37] to measure HR and RR. The system exposes the subject to a low-intensity microwave signal that is reflected back to the radar's sensor. Since respiration causes motion on the chest wall and abdomen, it effects on the reflected signal permits the deduction of the RR.

Iyer et al. [38] employed a Doppler radar-based radio frequency sensor for a smart home application to detect human subjects in a room by finding their RR and HR. Their proposed sensor simultaneously operates at 2.44 and 5.25 GHz frequency bands to transmit and receive the

generated signal. Based on the normal range of the HR and RR, they apply a bandpass filter on the received signal to eliminate noise. For their experiment, they collected a sample of 10 minutes of a 33-year-old male subject stationed within a distance of 1-m from the sensor. Subsequently, they used 'morlet' as the mother wavelet to decompose the spectrum of the obtained signal. Therefore, the maximum peak of the decomposed signal corresponds to the desired heart or respiration rate. Their experimental results demonstrate the applicability of this method to smart home applications.

Uenoyama et al. [39] used a 1,215 MHz-microwave radar antenna box with a maximum output power of 70 mW and two micro stripe antennas for radiating and receiving to measure the RR of sleeping subjects. The radar system was attached to the room ceiling, 200 cm from the clothed healthy subjects laying in a bed. They considered the subject's chest as the ROI to track the respiration-induced movements. They measured the RR by calculating the Fast Fourier Transform (FFT) of the chest wall motions measured by the radar. The consider the peak of the FFT signal to reflect the RR.

#### 2.1.2.2 Laser

Kondo et al. [40] proposed an RR measurement scheme that use a laser diode to measure the distance between the chest wall and the sensor. By plotting the changes in the latter distance against time, they obtained respiratory waveforms that led to the RR estimation.

Min et al. [41] used an ultrasonic proximity sensor to measure the chest-wall motion. However, with such method, the texture of the subject's clothes can affect the RR measurement accuracy [42].

#### 2.1.2.3 Thermal Camera

Using thermal cameras creates new opportunities to realize non-contact vital signs measurement techniques. As exhaled air is coming from inside the lungs, it has a higher temperature than inhaled air. Yang et al. [43] used a thermal camera to measure HR and RR by

monitoring periodic temperature changes that result in infrared light modulation near superficial blood vessels and the nasal area. Murthy et al. [44] used a cooled mid-wave infrared Phoenix camera with thermal sensitivity of 0.0250C to record thermal video sequences and a second-order statistical method to measure the contact-free RR of subjects. They conducted their experiment in a dimly lit room to avoid problems with reflection. They selected an area at distances ranging from 6-8 feet from the nose tip as the ROI that shows the differences in temperature during breathing. They used an adaptive thresholding algorithm to distinguish the skin of the face from the no-skin area in the profile view of the subject. They divided their process into two phases, training and testing. In the training phase, they trained the system on the subject's breathing temperature profile. Hence, they monitored the temperature of the ROI during the first breathing cycles and computed the Tmin (minimum temperature) Tmax (maximum temperature) and the average of Tmin and Tmax (Tave). They considered that the temperature distribution in inspiration is between Tmin and Tave and for expiration is within Tave and Tmax. After repeating this process for several breathing cycles, they computed the mean and variance of these distributions (they considered a normal distribution). In the testing phase, by considering the means and variances of temperature distributions in training phase, if the difference between the incoming temperature distribution to the distribution of the temperature for inhalation in training is lower, the pixels of the ROI containing incoming temperature are labeled as inspiration and otherwise, they are labeled as expiration. Based on this decision, they update the parameters (weight, mean and variance) of the inspiration and expiration distributions. After detecting the first full cycle of breathing, they can estimate RR with the time stamps of all the frames in the cycle. However, this approach can yield inaccurate results due to the variation in room temperature [45].

To compare the microwave radar-based method with thermal imaging, we considered the Uenoyama et al. [39] and Murthy et al. [44] as representative studies. For the thermal imaging

method presented in [44], they performed testing on ten participants. They did not report the age, health condition, and body temperature of the participants. They did not specify the room temperature where the experiment took place. They did not investigate whether the accuracy of the method would be affected by a rise or drop in body or room temperature. For instance, they did not explore the effect of fever on the correctness of their measurements. For the radar-based method [39], they conducted testing on sixteen subjects. Eight of the subjects were healthy and young (mean age, 25 years; range, 21–44 years); the other eight were elderly individuals (mean age, 69 years; range, 66– 75 years) with some disorders (hypertension 4/8, angina pectoris 2/8, aortic regurgitation 1/8, left upper lobectomy 1/8, prostatauxe 1/8, gastric ulceration 1/8). This variety in the subjects' characteristics highlights the comprehensiveness of their investigation. However, they did not test the feasibility of their method for situations where the subject is not laying in the bed on their back (i.e. sleeping on their side or stomach). Moreover, they did not explore the effects of covering the subjects with blankets. For the experiments of both studies, the subjects were limited in their natural movements.

#### 2.1.2.4 Depth Camera

Wijenayake et al. [46] employed a depth camera to acquire the depth data of the chest respiratory motion. This method requires the attachment of dot markers onto the person's chest area. Moreover, any wrinkles on the subject's garment within the chest wall area can have adverse effects on the accuracy of the results [46].

#### 2.1.2.5 RGB Camera

Using an RGB camera, we can measure the RR using two methods: 1) monitoring the motion of an ROI that reflects breathing, such as the chest, abdomen, shoulders, neck, or nostrils [8]; or 2) extracting the RR from an rPPG signal obtained from a video.

Janassen et al. [8] observed that respiration induces subtle trunk motions on chest or abdomen

which can be captured by dense optical flow with a camera to extract the respiratory signal. They employed the dense optical flow algorithm to estimate pixel motion vectors. They observed that their methods' performance degrades at farther distances from camera and under varying lighting conditions.

Lin et al. [47] and Tran et al. [48] used a standard camera to record changes of simple harmonic movements during respiration to estimate RR. In [47], they detect the face to identify the relative upper body location. Then, they model the movements of the upper body induced by respiration from the vertical variation of the optical flow to calculate the RR. In [48], they localize the chest region through body segmentation and by tracking respiration-derived optical flows of the upper body, they calculate the average breathing signal.

Prathosh et al. [42] and Zhao et al. [49] investigated imaging changes in the reflected light caused by respiration-induced motion to measure RR. In [42], they used an RGB camera while in [49] they used a near-infrared camera that is sensitive to light in the visible and near infrared region. Their base principle is that respiratory movements of the body surface shapes the path length of light reflect off the chest wall. Hence changes in the reflected light are correlated with respiratory events which leads to RR estimation.

Shao et al. [50] estimated RR by monitoring a region around the edge of the shoulder and implementing a differential detection and motion-tracking algorithm. The accuracy of the algorithm was negatively affected by subject movements.

Al-Naji et al. [51] used a motion magnification technique based on wavelet decomposition and an elliptic filter to magnify breathing movement of the chest caused by inhalation and exhalation. After magnification, they implemented frame subtraction to detect the motion and measure RR. Wei et al. [52] proposed monitoring the motion within two ROIs: the face and neck. They applied a blind identification algorithm to obtain a respiratory motion artifact. Then, they leveraged a

kurtosis-based method to automatically select the best signal carrying RR information among all decomposed signals.

Jorge et al. [53] proposed a convolutional neural networks based method for skin segmentation. To extract RR, they implement ICA as a blind source separation technique for a linear combination of multiple shape measurements over the exposed skin region. Also, Yang et al. [54] demonstrated the feasibility of RR monitoring based on nostril movements using input head-up facial videos. They measured the nose position using the face parts detection algorithm for each frame in a target video. By measuring the horizontal width and vertical length of the nostrils in each frame, they monitor the nostrils movements within each 1-minute video to estimate RR. Wu et al. [55] developed the Eulerian Video Magnification(EVM) technique to amplify respiratory-induced motions in the videos of subjects to monitor RR. Alghoul et al. [56] compared the EVM and ICA approaches. Their investigation revealed that ICA-based methods generally perform better than EVM-based techniques.

The main challenge associated with all of the aforementioned motion-based techniques is the differentiation between respiratory-induced movement and other unrelated motions.

The blood absorbs more light compared to the surrounding tissue. Hence, we can obtain a PPG signal by assessing the effects of the variation in the blood volume on the light reflected from the skin. Therefore, it affects the reflectance of the light [57].

Earlier work has shown that these variations can also be measured at a distance to obtain rPPG [58]. Furthermore, previous studies have proven the feasibility of using a regular RGB video camera for rPPG monitoring in ambient light conditions [59], [60].

Respiration modulates the PPG in three ways [58]: baseline, amplitude, and frequency modulation. Respiration causes pressure variations in the chest-abdominal area, and leads to pressure changes in the blood vessels and blood volume. Karlen et al. [61] estimated the RR from

PPG by averaging three independent RR values obtained from each of the respiratory-induced variations. Moreover, Gastel et al. [62] used the linear combination of the normalized color channels that provides rPPG and extracted the respiratory-induced baseline modulation to obtain the RR. Their algorithm detects the face and divides it into sub-regions. They calculate the Signal to Noise Ratio (SNR) for each sub-region. The sub-region with the highest SNR is used for RR estimation. However, this approach is noise-sensitive and fails to estimate RR from noisy rPPG signals [63]. Similarly, Tarassenko et al. [64] measured low-frequency amplitude variations of the camera reflectance signal to obtain RR. Using 13 facial features in their implemented face registration algorithm, they recognized faces in the frontal and profile view. Once the patient's face is recognized in every video frame, they utilized a non-parametric Bayesian image segmentation algorithm to segment the image into three categories: face, upper body, and background. Then, they averaged the RGB channels derived from the face region to fit their auto-regressive model. They implement this model and a pole cancellation scheme to construct an appropriate map of the spatial distribution of RR from the coefficients of the auto-regressive model.

Poh et al. [6] averaged the color intensity of RGB channels of ROI. Afterwards, they used ICA [59] as a blind source separation technique to extract the desired informative channel containing blood volume pulse information as PPG signal from a linear mixture of the underlying signals. Afterward, they determined RR from the power spectrum of quantified Heart Rate Variability (HRV). In another study by the same authors, they selected the second component produced by ICA as the PPG signal [57]. Although RR estimation from HRV is effective for healthy young subjects, this method might not provide accurate results for elderly subjects with chronic diseases [65]. Sanyal and Nundy [7] proposed a Hue-based (from the HSV color space) rPPG measurement technique. They averaged the time dependent fluctuations of Hue to measure arterial pulsations

and eventually RR from the forehead region during limited user motion.

***Table 1: Summary of the measurement set up of the main references works and the proposed method***

| Ref. | Activity | # of Subjects | Region of Interest and Detection Algorithm |
|---|---|---|---|
| [62] | 1: Seated Stationary<br>2: Supine Position | (0 F, 4M)<br>(2 neonates) | Face: Manually initialized |
| [64] | Supine Position | 46<br>(10 F, 36 M) | Face: Face registration algorithms of Everingham et al .[66] |
| [6] | Seated Stationary | 12<br>(4 F, 8 M) | Face: OpenCV Library (a boosted cascade classifier)[67] |
| [7] | Seated stationary | 25<br>(10 F, 15 M) | Forehead: OpenCV Library (a boosted cascade classifier)[68] |
| **Proposed Method** | 1: Seated Stationary<br>2: Walking freely (Distance 5 meter from the camera) | 10<br>(3 F, 7M) | Forehead: OpenCV Library (a boosted cascade classifier) [67]<br>KLT feature Tracker[69] |

Table 1 compares the existing related rPPG based RR estimation algorithms with the proposed method in terms of the activity of the subjects, the number of subjects, and applied algorithms for face and/or forehead detection. The activity includes sitting without motion, supine position without motion and movement without restrictions. Also, the gender of the participants is reported in the number of subject's column. Table 2 depicts the color spaces in use and source separation schemes with filters and their related noise reduction methods. The last column describes the component select after the application of source separation.

*Table 2: Summary of the algorithm of the reference works and the proposed method*

| Ref. | Color Space | Blind Source Separation Method | Noise Reduction | Component selection |
|---|---|---|---|---|
| [62] | RGB | N/A | A Band Pass Filter | No source separation applied. Linear combination of Red, Green and Blue |
| [64] | RGB | N/A | A Band Pass Filter | Average color intensity of Green channel |
| [6] | RGB | ICA | -Five Point Moving Average<br>-Hamming Window Filter | Component with the highest peak in PSD |
| [7] | HSV | N/A | A Band Pass Filter | No source separation applied. Hue Channel |
| **Proposed Method** | RGB | ICA | -Butterworth Band Pass filter<br>-CEEDMAN decomposition for noise removal | Machine learning method for the selection of PPG channel.<br>Machine learning method for the selection of the intrinsic mode function that reflects RR. |

We employ Poh. et al [6] and Sanyal and Nundy [7] as benchmarking methods. Hence, we investigate these studies in more detail and compare their results with the proposed method in chapter 4.

Fig. 1 & 2 depict the benchmarking methods proposed by Poh. et al [6] and Sanyal and Nundy [7] respectively. Both benchmarking methods use a face detection algorithm to localize the face of the stationary subjects. Poh et al. [6] used the center 60% width and full height of the face bounding box as the ROI for further calculations. They separated the ROI into RGB channels and averaged related pixels for each frame to form the raw signals. After detrending and normalizing the three acquired signals, they decomposed them into independent component signals using ICA as a blind source separation technique. They designated one of the ICA outputs as the most reflective of PPG. Poh. et al. [6] chose the ICA outputted signal with the maximum integrated power spectrum. However, Poh. et al. [57] heuristically chose the second component produced

through ICA. After filtering the chosen ICA output in the frequency range of interest, they calculated the power spectral density of the extracted HRV signal to estimate RR [57].

Ghanadian et al. [71], argued that we cannot predict a priori the order of ICA produced channels. Hence, they proposed a stacked Machine Learning (ML) algorithm to select the most informative channel at the ICA output [71].

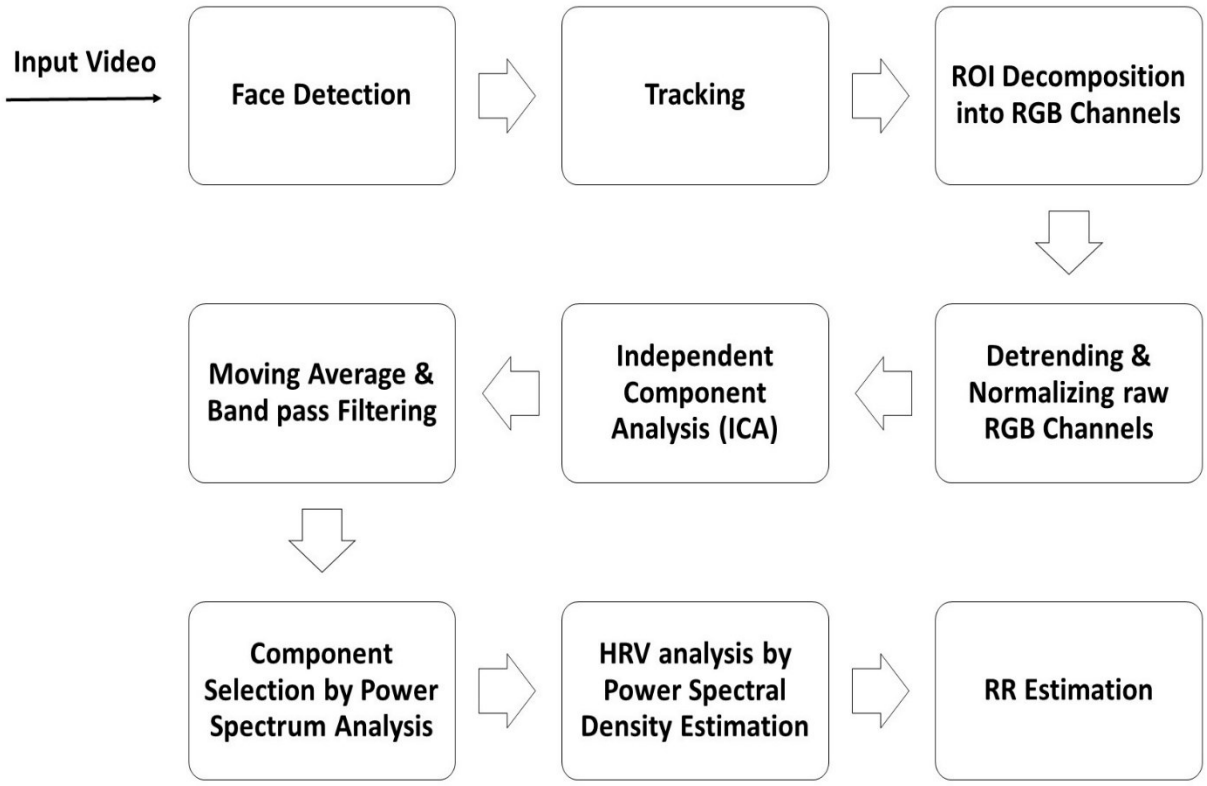



***Figure 1: RR prediction algorithm for Poh et al. [6] method***

Sanyal and Nundy [7] converted the RGB pixels to the corresponding HSV color space. As they tracked the forehead, they chose a specific range in the Hue that corresponds to the color of the skin in their ROI. They averaged the Hue of all pixels in each frame to obtain a raw PPG signal. Then, they applied a band pass filter in the frequency range of interest on the obtained PPG signal. They applied the FFT to detect the dominant frequency of the obtained PPG signal that leads to RR estimation.

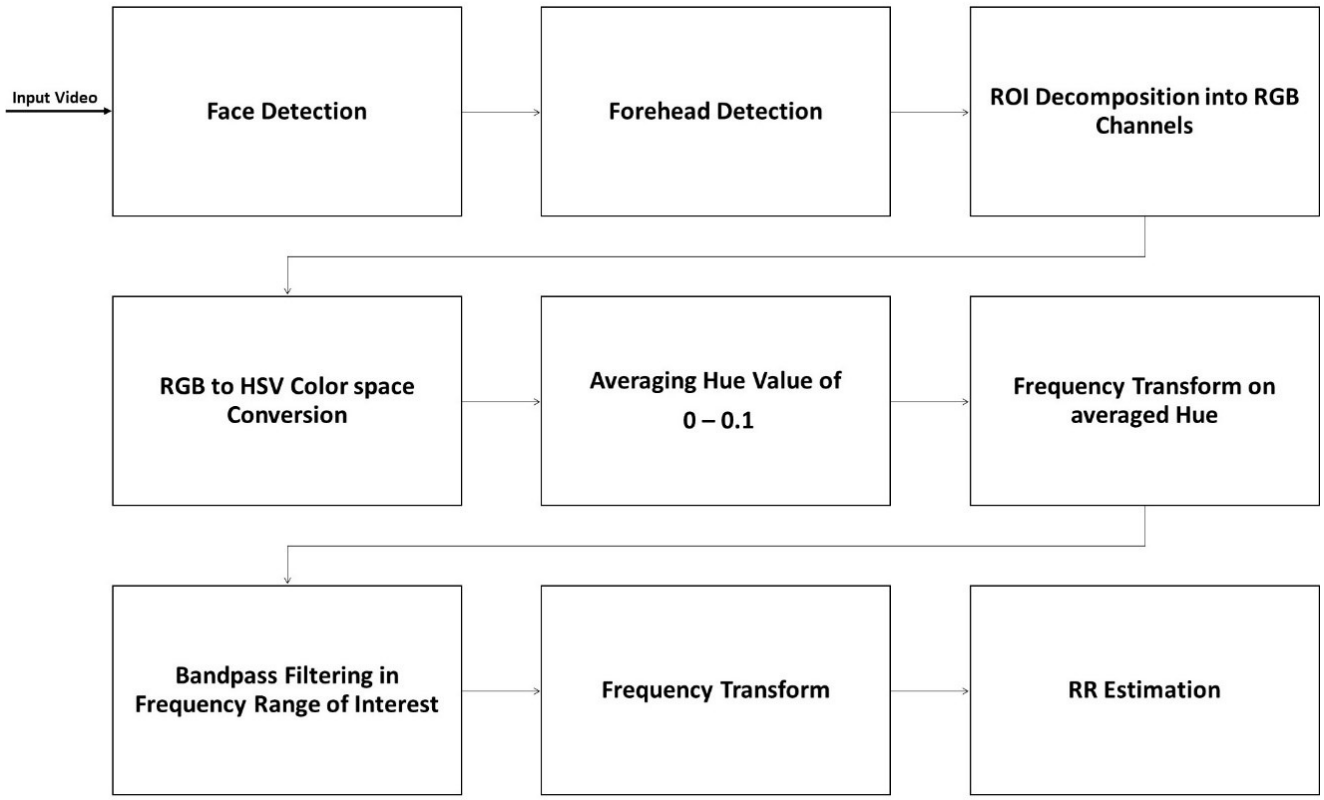



***Figure 2: RR prediction algorithm for Sanyal et al. [7] method***

## 2.2. The Region of Interest Detection Systems

To develop a remote RR measurement system using video, we must first detect the ROI. Based on the literature of rPPG based RR monitoring, the face is the most commonly used ROI.

### 2.2.1. Face Detection

Face detection is a mature research area with techniques applicable to a wide variety of systems, namely biometrics, indexing of images, human-computer interfaces and computer vision [72].

All face recognition algorithms can be categorized into two major parts: (1) face detection and normalization (2) face identification [73]. Face detection consists of the separation of the image

into two sections that are the face and the background. Although similarities exist between faces, they can vary in terms of the subject's age, skin color, and facial expression [74]. Hjemal et al. [75] divides the face detection techniques into two categories: feature-based and image-based techniques. The feature-based technique detects facial features to localize the face. However, the differences in environmental conditions and facial features between subjects present challenges when such technique is applied [74]. Image-based face detection employ pattern recognition methods that rely on machine learning approaches like support vector machine and neural networks [74].

Yuille et al. [76] detected facial features using deformable templates. The templates of features of interests such as lips or eyes, are described by parameterized templates. The parameterized templates provide a generalized description of the expected shape of the facial features. Feature detection is achieved by correlating facial segments with the templates. Brunelli et al. [77] uses a dataset containing facial images as reference to obtain facial features. Each unclassified input image is compared with the reference facial features dataset and the result of this comparison is a matching score. The image with the highest matching leads to face recognition. The method proposed by Jin et al. [78] employs a combination of template matching and color information. They used color information to segment eye-pair candidate and finally detect face by template matching. The shortcoming of aforementioned approach is the high computational procedure of template matching methodology. Also, the existence of similar skin color on background can cause false detection [79]. To overcome these deficiencies, Wang et al. [79] used a self-adaptive template to overcome the skin-like color of background. Afterwards, they adopted a secondary template matching to decrease the amount of calculation and increase the speed of matching process.

It has been proven that the combination of skin color segmentation, shape analysis and motion

information for locating and tracking the face [80,81,82]. In Hunke et al. [83] neural network-based method, they use a face color classifier to extract all areas containing face-like colors. Then, they feed the color and movement information obtained from the difference between pixels in consecutive frames to a neural network that can detect the face.

In our work, we detect the human faces and eyes using a Haar cascade scheme [68] (more details in chapter 3). Using the location of the eyes and face, we calculate the location of the forehead (i.e. our ROI).

## 2.3. Independent Component Analysis (ICA)

Independent component analysis (ICA) is a blind source separation technique that separates multivariate signals from underlying sources [6]. It is a computational method for distinguishing a multivariate signal into additive subcomponents.

Poh et al. [6], [57] used ICA to extract the desired informative channel containing blood volume pulse information from a linear mixture of the acquired RGB signals of recorded video. We employ this technique in the proposed method (see Fig.3). Through this process, ICA decomposes the recorded RGB signals to three independent source signals. One of these resulting signals typically better reflects the PPG information.

If we denote the intensity of the RGB channels in the ROI of the recorded video by $y$, then the intensity of the RGB channels at time $t$ is $y_1(t)$, $y_2(t)$ and $y_3(t)$, respectively. When they feed the color intensity signals to ICA, they obtain the output $x_1(t)$, $x_2(t)$ and $x_3(t)$. Usually, one of these signals reflect the measured PPG better than the others.

The input of the ICA is a linear mixture of the RGB derived underlying sources:

$$\boldsymbol{Y(t) = AX(t)} \qquad (1)$$

Where $\boldsymbol{y(t)}$ is vector $[\boldsymbol{y_1(t)}, \boldsymbol{y_2(t)}, \boldsymbol{y_3(t)}]^{\boldsymbol{T}}$ and $\boldsymbol{X(t)}$ is vector $[\boldsymbol{x_1(t)}, \boldsymbol{x_2(t)}, \boldsymbol{x_3(t)}]^{\boldsymbol{T}}$ and $A$ is a square matrix that contains mixture coefficients named $a_{ij}$. Therefore, to realize the relationship between $\boldsymbol{y(t)}$ and $\boldsymbol{X(t)}$, we have to calculate the inverse of matrix A denoted as $\boldsymbol{W}$. Once they calculate $W$, the source signals $\boldsymbol{X(t)}$ can be estimated.

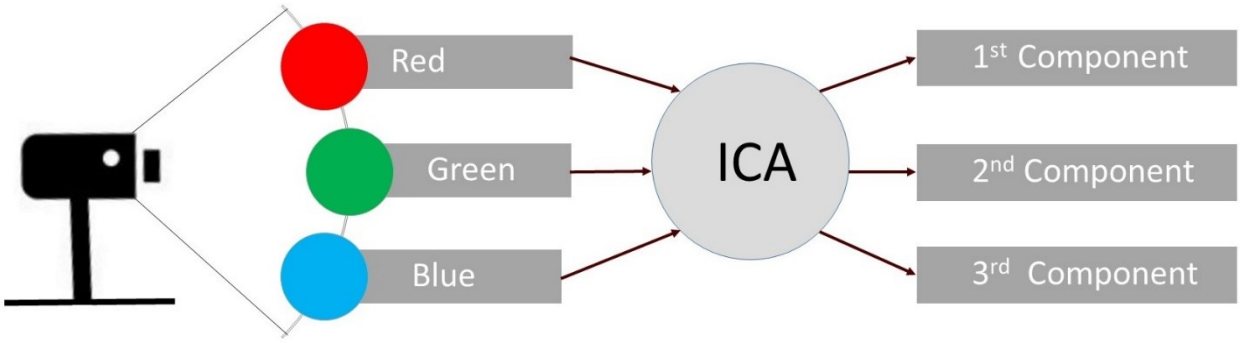



***Figure 3: Independent Component Analysis***

## 2.4. Signal Processing

It is not feasible to estimate RR from ICA outputted components as they require noise reduction. To the best of our knowledge, only Poh et al. [6,57] method use ICA to estimate RR from selected ICA-driven component. For noise reduction, they utilize five-point moving average and hamming window filter. Moving average filter is a linear-phase finite-impulse-response filter that is easy to implement [84]. This filter is applied to data points of the signal by averaging various subsets of the signal. It is defined as follows:

$$\text{(for five points)} \quad Z(n) = \frac{z(n-2)+z(n-1)+z(n)+z(n+1)+z(n+2)}{5} \qquad (2)$$

They are generally used to remove motion artifacts and works well for intermittent noise [85].

Hamming window filter is a type of band pass filter that reduces the ripples in the signal. It can provide a more accurate representation of the original signal's frequency spectrum [86],[87].

## 2.5. Empirical Mode Decomposition

Empirical mode decomposition (EMD) [88] is a technique that has been successfully implemented for artifacts removal in the PPG signal [89]. It is an adaptive decomposition method that derives its required basis function from the signal itself [88]. Fig.4 depicts the flow chart of EMD methodology. It decomposes the given signal into its Intrinsic Mode Functions (IMF). To define a component as IMF, it should satisfy two main conditions:

1. The number of extreme values and the number of zero crossings should be equal or differ by no more than one.
2. The average of local minimums and maximums should be zero.

Obtaining IMFs requires a process called sifting. A signal $x(t)$ is divided into details, $d(t)$, and the residual, $m(t)$. Therefore, $x(t)= m(t)+d(t)$. By considering $d(t)$ as the first IMF, the sifting process is repeated on the $m(t)$, $m(t)=x(t)-d(t)$. Therefore, to find the first IMF of the given signal $x(t)$, we should iterate the loop as follows:

1. Find the local minimum and maximum of $x(t)$.
2. Compute the maximum and minimum envelop of $x(t)$ to calculate their average.
3. Find the $a(t) = x(t)-m(t)$ to check whether $a(t)$ is an IMF, and if it is not, assign $a(t)$ to $x(t)$ and go back to step 1.

Madhav et al. [90] and Grade at al. [91] utilized EMD to estimate RR by decomposing the PPG signal. Their results revealed that EMD can be efficiently used to extract RR from PPG signals.

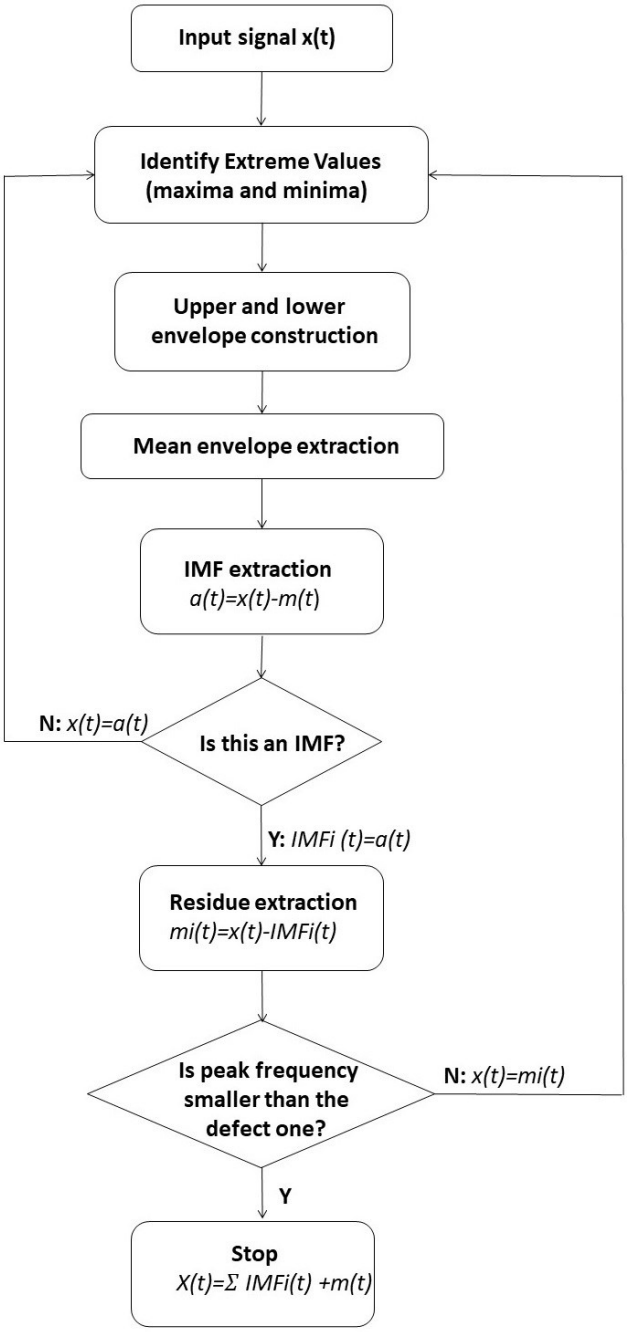



***Figure 4: Empirical Mode Decomposition Flow Chart***

## 2.6. Ensemble Empirical Mode Decomposition

EMD presents some shortcomings associated with the presence of oscillations of very disparate amplitudes in a mode, or the presence of very similar oscillations in different modes, known as "mode mixing" [13]. Hence, ensemble empirical mode decomposition (EEMD) was utilized to

eliminate the mode-mixing effect [92].

The steps for the EEMD algorithm are as follows:

1. Consider signal *x(t)* as the sum of the target signal *z(t)* and white noise *n(t)*: $x(t)=z(t)+n(t)$
2. Decompose *x(t)* using the EMD algorithm described in section 2.5.
3. Repeat the loop from step 1 and 2 by adding different white noise at each time. The new IMF combination that we will obtain from each iteration is $R_{ij}$ where *i* is the iteration number and *j* is the IMF scale.
4. For the final output, obtain the ensemble of the final IMF as EEMD $R_j(t)=\sum_{j=1}^{zi} R_{ij}(t)$ where *zi* denotes the trial numbers.

Motin et al. [93], utilized EEMD to decompose the PPG signal to extract HR and RR. They applied the FFT to extract the frequency at the maximum power for each generated IMF. After determining the dominant frequencies, they selected the IMFs with dominant frequency in the RR range for further processing. Then, they applied PCA on the selected IMFs to produce their Principal Components (PCs). Therefore, the first PC has the most variation in selected IMF which represents RR. For RR estimation, they applied FFT on the first PC to determine the frequency containing the maximum power and multiplied by 60.

In the Orphanidou et al. [94] study, they developed an algorithm for estimating RR from corrupted and noisy PPG of ambulatory signals. Therefore, they used EEMD to obtain clear IMFs that might contain RR information.

Ambekar et al. [95] utilized EEMD on the PPG signal to obtain the IMFs.Afterwards, they singled out IMFs which are in the frequency range of the RR. Then, they summed all selected IMFs to obtain a single signal, which they use to compute the RR. Moreover, they showed significant improvement in their results obtained by EEMD over EMD.

## 2.7. Machine Learning

In the proposed solution, we utilize a Machine Learning (ML) algorithm that uses specific characteristics of the IMFs in both time and frequency domain to detect the best IMF.

ML enables the computers to learn using statistical techniques [96], [97]. Through this procedure and using mathematical models, the machine learns from a given dataset known as "training data" to make prediction or decisions. Fig.5 shows the concept in different types of ML.

The type of ML algorithms are conventionally categorized based on their learning approaches as:

- Supervised Learning: The algorithms use labeled data with known input and output data. Based on the data, the machine recognizes the relationship between the input, labels, and output, to predict the label of new inputs.
- Unsupervised Learning: The algorithms process an unlabeled input dataset during training and learn its structure by clustering related data points.

In addition to these categories of learning, ML models can be trained through a process of reinforcement learning. In this category, the agent learns the behavior through trial-and-error interactions with a dynamic environment [98].

A supervised learning algorithm can be further categorized as a classifier or regressor.

### 2.7.1. Classification

Classification modeling is the task of identifying to which class a new input belongs based on the training dataset which has known memberships or labels. Hence, the output of the classifier is a label that describes the class of the input.

### 2.7.2. Regression

Regressing is a statistical process of defining a function from input variables to a continuous output variable. Therefore, the output of regression is a continuous value.

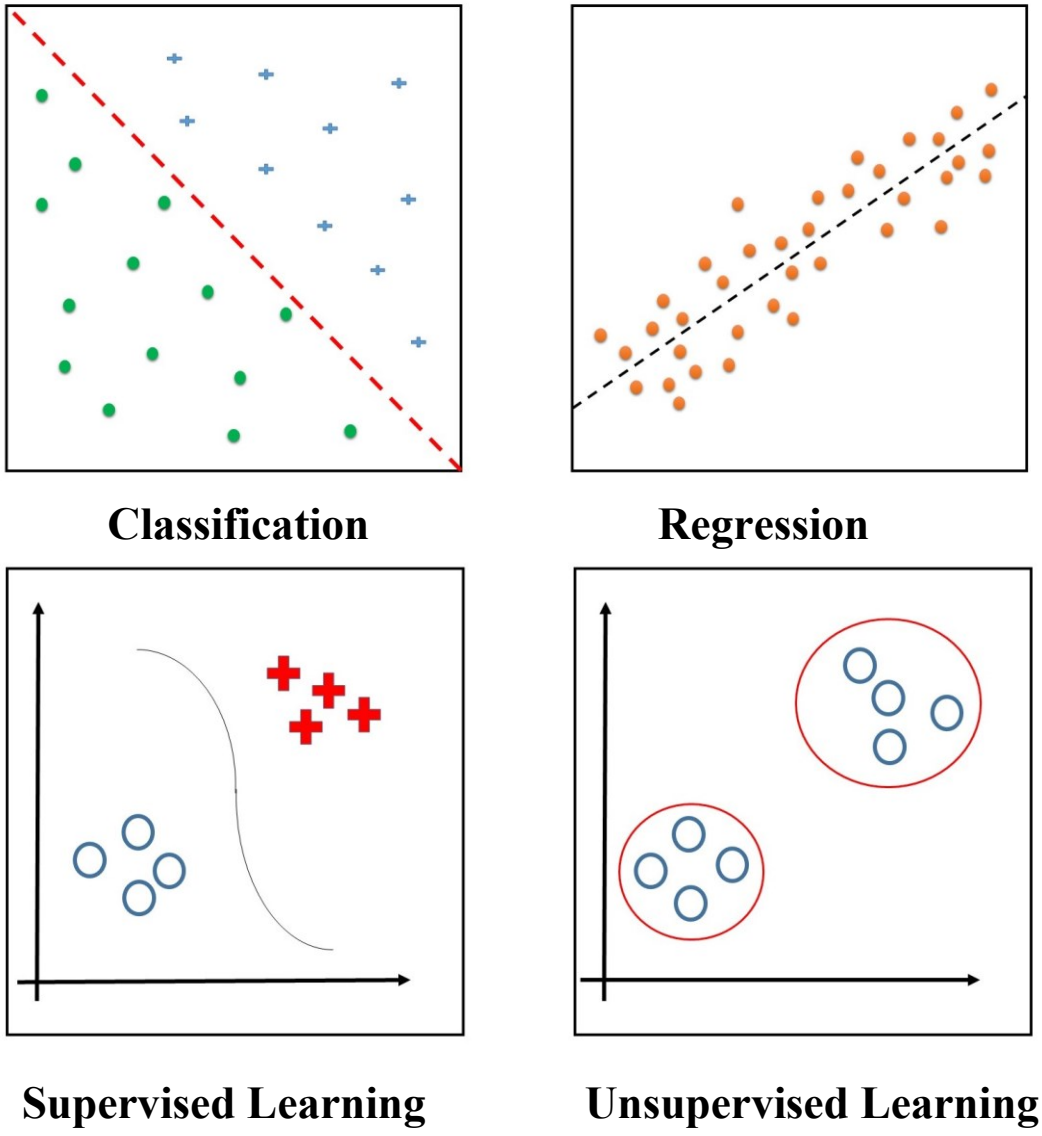



*Figure 5: Machine Learning categories*

## 2.8. Summary of Chapter 2

In this chapter, provided a brief explanation of RR and the importance of its measurement in both contact based and non-contact based methods.

We described how an RGB camera can be leveraged to the RR using two methods:

1) Monitoring the motion of an ROI that reflects breathing, such as the chest, abdomen, shoulders, neck, or nostrils [8].

2) Extracting the RR from an rPPG signal obtained from facial videos.

Respiration modulates the PPG in three ways [58]: baseline, amplitude, and frequency modulation. The blood absorbs more light compared to the surrounding tissue. Hence, we can obtain a PPG signal by assessing the effects of the variation in the blood volume on the light reflected from the skin Furthermore, previous studies have proven the feasibility of using a regular RGB video camera for rPPG monitoring in ambient light conditions [59], [60]. Table.1 & 2 summarize and compare the experimental setups and algorithms of previous rPPG methods.

We based our rPPG approach on Sanyal and Nundy [7] and Poh. et al [6] method. Fig.2 & 1 depicts their procedure of their approach respectively. They chose face as the ROI and since this region is the most commonly used ROI in the literature, we selected the same ROI.

# Chapter 3. Proposed Method

Fig. 6 &7 depicts our approach which we base on the benchmarking methods proposed by Poh. et al. [6,57] and Sanyal and Nundy [7] respectively.

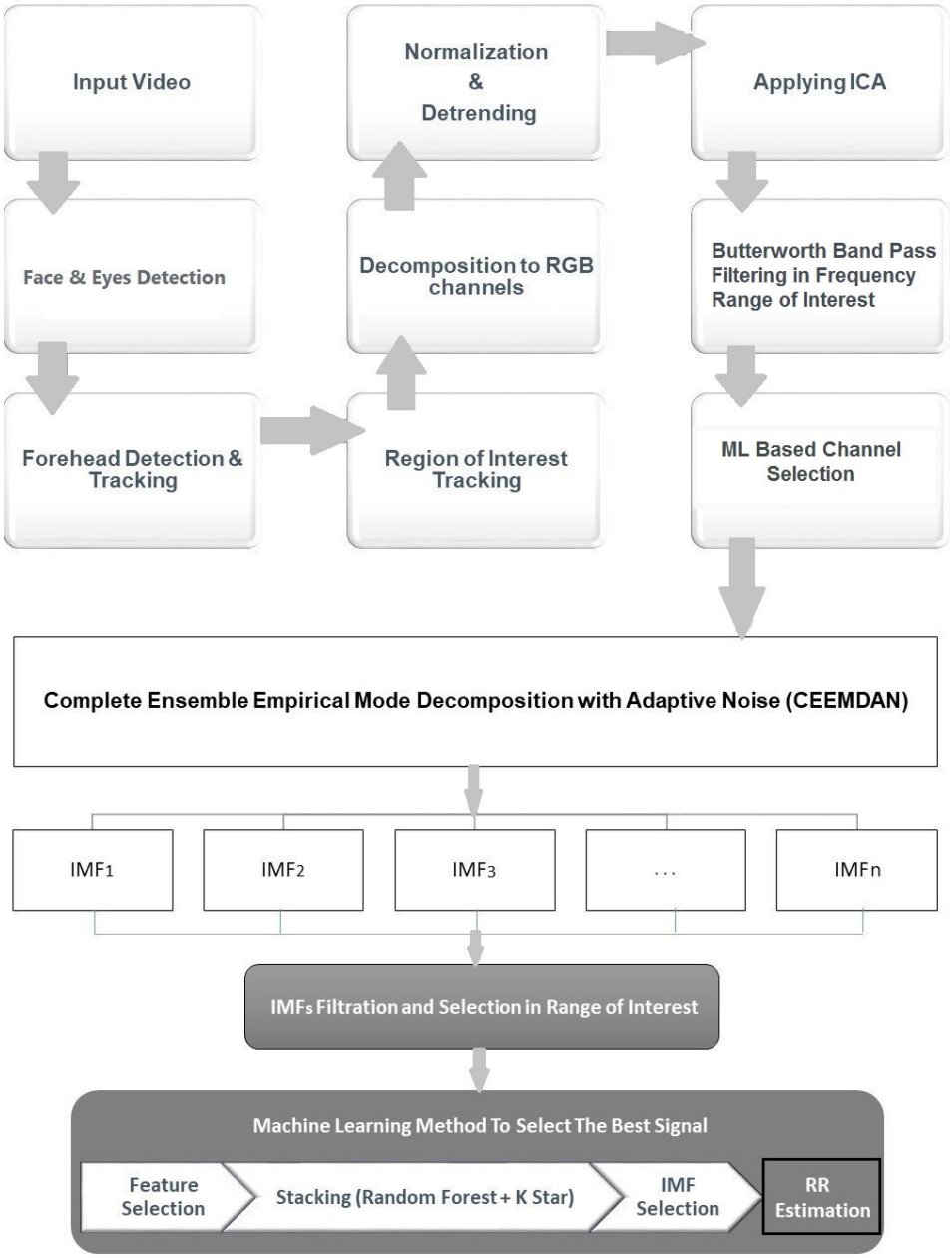



***Figure 6: RR estimation algorithm of the proposed method***

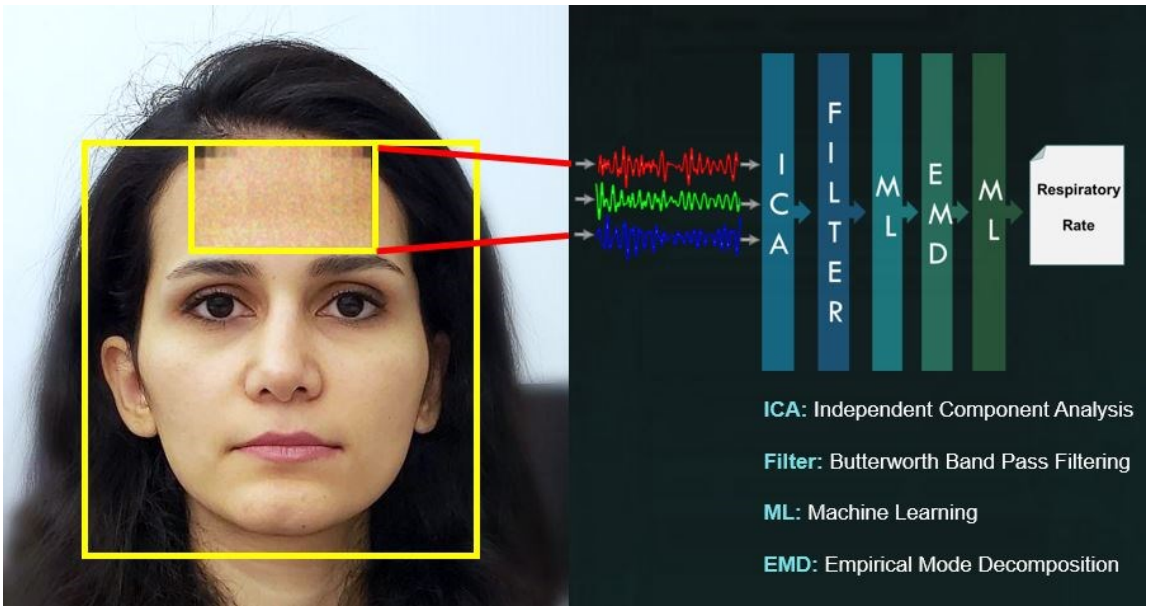



***Figure 7: RR estimation procedure from the forehead***

## 3.1. Forehead Detection and Tracking

According to Fallet et al. [99], the forehead region contains the largest power at the HR frequency and thus can allow us to obtain an informative PPG signal. Hence, we detect and track the forehead as our desired ROI.

In the first frame, a Haar cascade scheme detects the face and eyes using OpenCV [68]. This is a machine learning method trained from negative and positive images and is used for various object detection applications, in particular, for the detection of the eyes and face [67].

We detect the forehead based on the Sanyal and Nundy [7] approach. The Haar cascade function returns two bounding boxes representing the eyes and face respectively. Hence, it returns the coordinates of the top left corner (X and Y), height (H) and width (W) of each box (3,4).

Box of eyes =[Xeye,Yeye,Heye,Weye] *(3)*

Box of face= [Xface,Yface,Hface,Wface] *(4)*

With the aforementioned parameters of the eyes and face boxes, we calculate the forehead's bounding box using (5).

Box of forehead = [Xeye + (Weye*0.45),Yface, Weye* 3.2,(Yeye - Yface)*0.8] *(5)*

Using the strategy described in [6,57], we optimize the parameters to maximize the ROI information without capturing unwanted facial features (e.g. eyebrows or sideburns). This would increase the probability of obtaining an accurate signal as the subject moves away from the camera. (See Fig. 8).

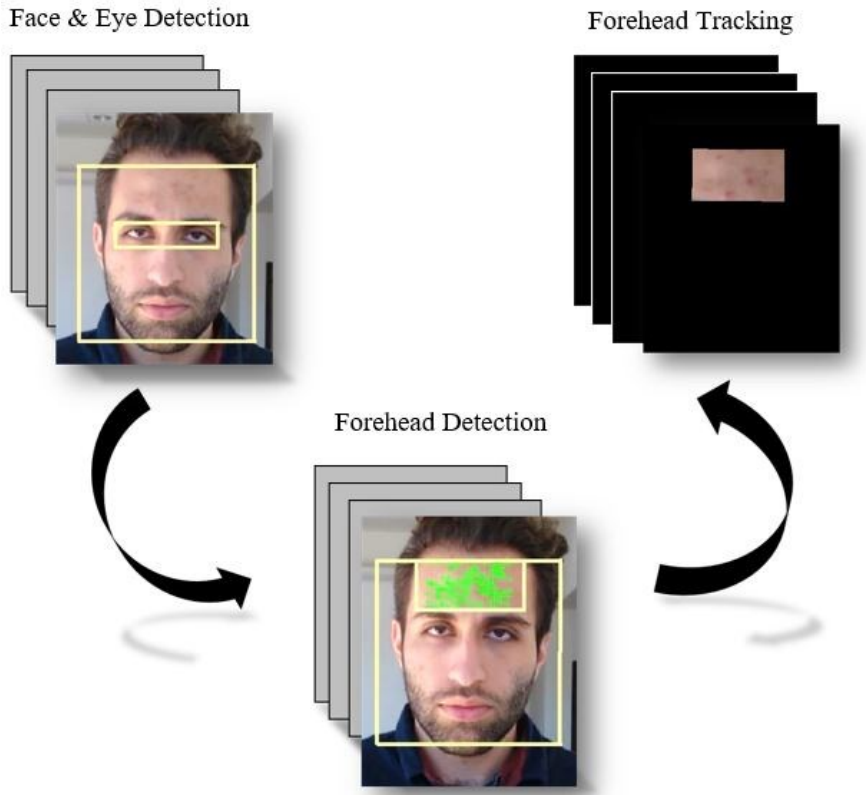



***Figure. 8: Forehead detection and tracking***

After detecting the forehead, we track it in subsequent frames using the Kanade–Lucas–Tomasi (KLT) algorithm [69] as the subject might be in motion, which causes the ROI's location to vary

significantly between frames.

## 3.2. Decomposition

We spatially average the pixels of the ROI for each of the RGB channels in each frame using the method presented in [6]. Using the smoothness priors approach [100] and a smoothness parameter of λ=2000, we de-trend the signal. Then, to normalize it, we subtract the mean of the signal and divide by its standard deviation. This is followed by the application of an IIR band pass filter of the 5th order corresponding to the frequency range of interest associated with RR (cut-offs: 0.05 to 0.75 Hz) that covers 3-45 breaths per minute. These processed RGB signals are presented to ICA to distinguish between the desired informative signal and noise. As one of the three ICA outputted signals should be selected for further processing, we apply the ML model proposed by Ghanadian et al. [71] to select the best ICA-based decomposed component that corresponds to the rPPG signal.

Next, the ML model derived output of ICA is presented to CEEMDAN to decompose the target signal, which leads to the RR estimation. CEEMDAN was proven as an important improvement on the Ensemble Empirical Mode Decomposition (EEMD) [92], which in turn is an enhancement on the Empirical Mode Decomposition (EMD) [88]. EMD is a data driven approach that can iteratively decompose a given signal. Using the dyadic filter bank behavior of EMD, the original signal can be expressed as a sum of amplitude and frequency modulated (AM–FM) functions called Intrinsic Mode Functions (IMFs) [101]. EMD can decompose the given signal *x(t)* into *n* number of IMFs. The input signal can be reconstructed from the IMFs and residue (rn) as follows (6):

$$x(t) = \sum_{i=0}^{n} IMF_i + r_n(t) \qquad (6)$$

To be considered an IMF, a signal must satisfy two conditions: first, the number of extrema and zero crossings must be equal or differ by one at most; second, the mean value of the upper and lower envelope must be zero everywhere.

However, EMD presents some shortcomings associated with the presence of oscillations of very disparate amplitudes in a mode, or the presence of very similar oscillations in different modes, known as "mode mixing" [13].

EEMD was proposed to overcome the aforementioned EMD issues. It performs the EMD over an ensemble of the signal with injected Gaussian white noise. Although the addition of white noise solves the mode mixing problem, the EEMD derived IMFs would be polluted by the added noise which cannot be completely canceled after a finite average. Therefore, the reconstructed signal from the obtained IMFs will differ significantly from the original signal. Moreover, the added noise may produce an incorrect number of modes.

In an attempt to overcome the problems associated with EEMD, Torres et al. [13] proposed CEEMDAN. By adding adaptive white Gaussian noise in every stage of the EMD, every IMF can be obtained by calculating the unique residue which effectively resolves the mode-mixing problem and results in a near zero reconstruction error. According to the principle of CEEMDAN, if we assume that Ej(.) is an operator that produces the j-th mode obtained by EMD, set zi(t) to be a zero mean unit variance white noise, and define $\beta i$ coefficients to select the signal to noise ratio at each stage, our signal x(t) can be decomposed as follows:

1. For $i$ = 1, …, L decompose each $x_i(t)=x(t)+\beta_0 z_i(t)$ using EMD (until we obtain its first mode) to compute the first IMF decomposed by CEEMDAN:

   $$\widetilde{IMF_1} = \frac{1}{L}\sum_{i=1}^{L} IMF_{i1} \qquad (7)$$

   At j=1, calculate the first residue:

   $$r_1(t) = x(t) - \widetilde{IMF_1} \qquad (8)$$

2. Using EMD and until obtaining its first mode, decompose $r_1(t)+\beta_1 E_1(z_i(t))$ for every $i$ = 1, …, L to calculate the second IMF of CEEMDAN:

   $$\widetilde{IMF_2} = \frac{1}{L}\sum_{i=1}^{L} E_1(r_1(t) + \beta_1 E_1(Z_i(t))) \qquad (9)$$

3. For $j = 2, 3, \ldots, C$, calculate the $j$-th residue:

$$r_j(t) = r_{j-1}(t) - \widetilde{IMF_j} \quad (10)$$

4. Using EMD, decompose $r_j(t) + \beta_j E_j(z_j(t))$ to obtain its first mode and calculate the next IMF of CEEMDAN (for every $i = 1, \ldots, L$):

$$\widetilde{IMF_{j+1}} = \frac{1}{L}\sum_{i=1}^{L} E_1(r_j(t) + \beta_j E_j(Z_i(t))) \quad (11)$$

5. Afterwards, we repeat the steps of 3 and 4 to obtain further IMFs of CEEMDAN until it satisfies the IMF criteria (i.e. the number of extrema and zero crossings must be equal or differ by one at most and the mean value of the upper and lower envelope is zero everywhere). Hence, the final residue value (12) and our signal (13) can be defined as: (C refers to the total number of IMFs)

6.

$$r_C(t) = x(t) - \sum_{j=1}^{C} \widetilde{IMF_j} \quad (12)$$

$$x(t) = \sum_{j=1}^{C} \widetilde{IMF_j} + r_C(t) \quad (13)$$

According to Colominas et al. [102], a value close to 0.2 for the standard deviation of the added noise effectively reduces the residual noise in the reconstructed signal and minimizes the reconstruction error. Consequently, we adopt this value.

## 3.2. IMFs Filtration and Selection

Using an ensemble size of 100, we obtain the IMFs using CEEMDAN. Afterward, we identify the IMFs containing artifacts and discard them. The normal range of RR for children aged 2 to 18 years' old and young adults is typically between 8 to 45 Breaths Per Minute (BPM) [91], [103]. The range of RR for older adults is within the mentioned range, although it is more restricted [63]. Hence, to distinguish between IMFs that contain no useful information and ones that may reflect

RR, we apply the FFT on each IMF to determine its dominant frequency (the frequency associated with the maximum power). Then we reject the IMFs with a dominant frequency that does not fall in the RR range. In particular, we consider IMFs with a dominant frequency less than 0.05 Hz (3 BPM) and greater than 0.75 Hz (45 BPM) as noise artifacts. We select IMFs with a dominant frequency within the range of 0.05–0.75Hz for additional analysis.

For instance, in Fig.3, the selected output of ICA (using the Ghanadian et al. [71] ML model) is decomposed to its IMFs. After dropping the IMFs with dominant frequencies outside the previously specified range, the first three maintained IMFs are selected. In the next step, we need to determine which one of these IMFs is most reflective of the RR information. We explain our procedure to achieve the latter in Section 3.3.

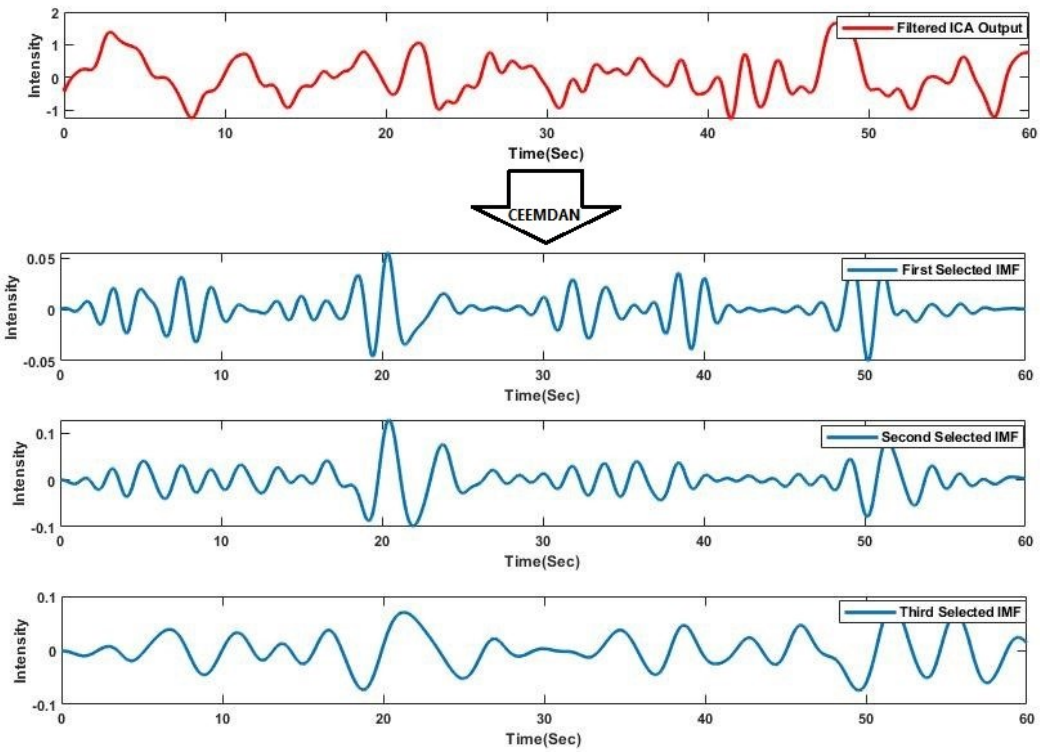



***Figure. 9: An example of decomposing selected ICA output with CEEMDAN***

## 3.3. IMF Selection

At this stage, we have to select the final IMF such that its dominant frequency represents RR. We utilize an ML algorithm that uses specific characteristics of the IMFs in both time and frequency domain to detect the best IMF. Initially, we selected thirty-five features from each IMF to train the ML algorithm. We ran the PCA (Principal Component Analysis) as a feature selection algorithm [104], to choose the most informative features. Below is a summary of the final selected features:

- We transform the IMF using a five-level decomposed Daubechies wavelet (order one) [105]. We apply FFT on the obtained wavelets and select the dominant frequency of each as features.
- We use the same wavelets of the previously described features. We select as a feature the ratio of the peak absolute magnitude of each wavelet to its Root-Mean-Square (RMS) value. We calculate the peak-magnitude-to-RMS ratio as follows:

$$peak\ to\ rms\ ratio = \frac{\|x\|_\infty}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}|x_n|^2}} \quad (14)$$

  Where $x$ is the given signal with $N$ values, and $\|x\|_\infty$ is the largest absolute value in $x$.
- We apply the FFT on the IMF and select the dominant frequency and its amplitude as features.
- We use the peak-magnitude-to-RMS ratio of the IMF as a feature.
- We detect the local maxima and minima points of the IMF, we calculate the average of these points and the difference between successive ones as features.
- We use the mean of autocorrelation with twenty lags as a feature. This feature measures the correlation between $xt$ and $xt + k$, *where* $k = 0..., K$ where $xt$ is a stochastic process. According to [106], the autocorrelation for lag k is: ($C_0$ is the sample variance of the time

series and T is sample size of x)

$$r_k = \frac{c_k}{c_0} \tag{15}$$

Where:
$$c_k = \frac{1}{T}\sum_{t=1}^{T-k}(x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{16}$$

- We compute the Zero-Crossing Rate (ZCR) as a feature. ZCR is the rate of sign-changes along a signal as defined by:

$$zcr = \frac{1}{T-1}\sum_{t=1}^{T-1} 1_{R<0}\,(x_t x_{t-1}) \tag{17}$$

Where *x* is a signal of length *T* and $1_{R<0}$ is an indicator for membership of an element in a given subset of the signal.

We conducted initial testing on several classifiers to determine the ones that render the best performance. Consequently, we identified the best three classifiers: Random Forest, K-star, and Rotation Forest. We present the summary of these classifiers' performance in Section 4.2.

Random Forest is an ensemble learning method that was developed by Breiman [107]. It contains multiple decision trees that train on the dataset and returns the average of their predictions. Random Forest models do not require pruning and are less likely to over fit [108]. Rodriguez et al. [109] proposed Rotation Forest as a classifier ensemble method with a built in feature selection mechanism. It uses the decision tree as a base classifier and splits the feature set into subsets. It runs PCA on each subset to obtain a new extracted feature set.

K-star [110] is an instance-based classifier that utilizes an entropic-based distance function to quantify the complexity of transforming one instance into another.

Moreover, we propose to utilize a stacking algorithm to improve performance. By stacking, we combine several classifiers, which leads to the emergence of a final, habitually more accurate classifier [111].

The ML model outputs “one” if it estimates that the IMF carries the RR information and “zero” otherwise. If the model produces more than a single “one” output for the considered IMFs (i.e. it judges that more than a single IMF carries the respiration signal), then we calculate the average of the RR comprised in all of these IMFs. Similarly, if the ML model deems that none of the IMFs carry respiration information, we calculate the mean RR that we can obtain from these IMFs.

## 3.4. RR Estimation

After selecting the IMF that contains the RR information, we transform it using FFT, extract the RR frequency ($f_{RR}$) by calculating the dominant frequency, and convert it to breath per minute (18).

$$RR = f_{RR} * 60 \ \ (Breath/\min) \tag{18}$$

## 3.5. Summary of Chapter 3

In this chapter, we explained our proposed method. We base our approach on the benchmarking methods proposed by Poh. et al. [6,57] and Sanyal and Nundy [7]. Since we detect and track the forehead as our desired ROI, we implement a Haar cascade scheme detects the face and eyes using OpenCV [68]. Based on the Sanyal and Nundy [7] approach, we detect the forehead using the detected bounding boxes of face and eyes. After face detection, we track the forehead in subsequent frames using the Kanade–Lucas–Tomasi (KLT) algorithm [69].

After spatially averaging the RGB channels of ROI, we smooth, de-trend, and normalize them respectively. This is followed by applying an IIR band pass filter of the 5th order corresponding to the frequency range of interest (3-45 breaths per minute) associated with RR (cut-offs: 0.05 to 0.75 Hz) .

The processed RGB signals are presented to ICA to distinguish between the desired informative signal and noise. we apply the ML model proposed by Ghanadian et al. [71] to select the best ICA-based decomposed component that corresponds to the rPPG signal.

Afterwards, the selected output of ICA is presented to CEEMDAN to decompose the target signal, which leads to extract the IMF that represents the RR. CEEMDAN was proven as an important improvement on the Ensemble Empirical Mode Decomposition (EEMD) [92], which in turn is an enhancement on the Empirical Mode Decomposition (EMD) [88].

After the application of CEEMDAN, we identify the IMFs containing artifacts and discard them. We consider IMFs with a dominant frequency less than 0.05 Hz (3 BPM) and greater than 0.75 Hz (45 BPM) as noise artifacts. Therefore, we ignore them and select the rest of the IMFs.

At this stage, we utilize an ML algorithm that uses specific characteristics of the IMFs in both time and frequency domain to detect the best IMF.

# Chapter 4. Experiment, Results, Evaluation and Discussion

## 4.1. Dataset

To collect our dataset, we recruited 10 consenting adult subjects (3 females and 7 males) with mean age of 27.3 and standard deviation of 2.5 years through the experimental setups shown in Fig.10. We recorded 250 one-minute-videos of subjects in stationary and movement modes as described in Sections 4.1.1 and 4.1.2, respectively.

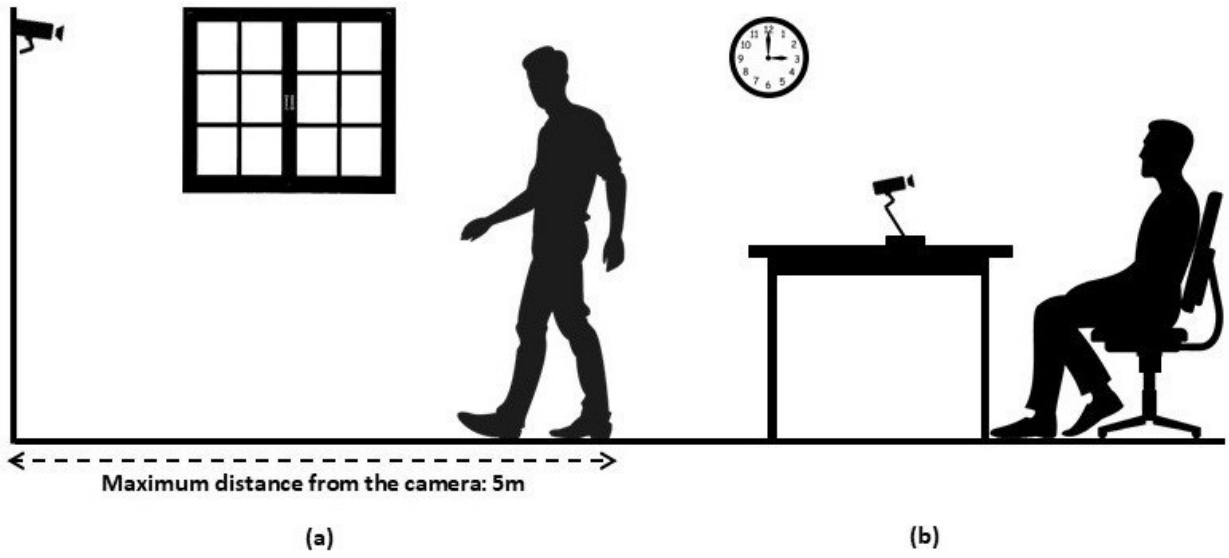



*Figure. 10: Experimental Setup for (a) Movement mode, (b) Stationary mode*

### 4.1.1. Stationary Mode

We instructed the subjects to wear a thoracic expansion sensor shown in in Fig.11 (Zephyr Bioharness[1]) that measures RR, based on respiratory-derived chest movements (sampling rate:

250Hz). Moreover, we provided them with a thumb counter (a mechanical device that incrementally counts in response to the pressing of its button). We instructed them to practice counting their breaths using the thumb counter. Once they felt comfortable with the breaths counting process, we asked them to sit in front of a webcam with minimum movements.


***Figure. 11: chest belt (Zephyr Bioharness)***

Once they felt comfortable with the breaths counting process, we asked them to sit in front of a webcam with minimum movements. We directed them to count their breaths using the thumb counter while we recorded a minute-long video of their face. We further corroborated the accuracy of their count by a visual inspection of their chest movements, using the method proposed by the Johns Hopkins University and John Hopkins Hospital [112,7].

Afterward, whenever we had a discrepancy of more than 5 breaths / minute between the sensor, thumb counter, or visual inspection, we discarded the video and collected another one. We adopted as ground truth the subject-measured RR recorded by the thumb counter. We repeated this process until we collected 5 videos for each subject, where each video had a length of 1 minute.

Fig. 10b & 12depicts our experimental setup for the stationary subjects. Our light sources consisted of a combination of ceiling and natural light coming through windows.

For video recording, we used the Logitech C270 RGB webcam recording at 30-frames per second and 640×480 pixels per frame. We mounted the camera on a tripod located one-meter far from the subject.

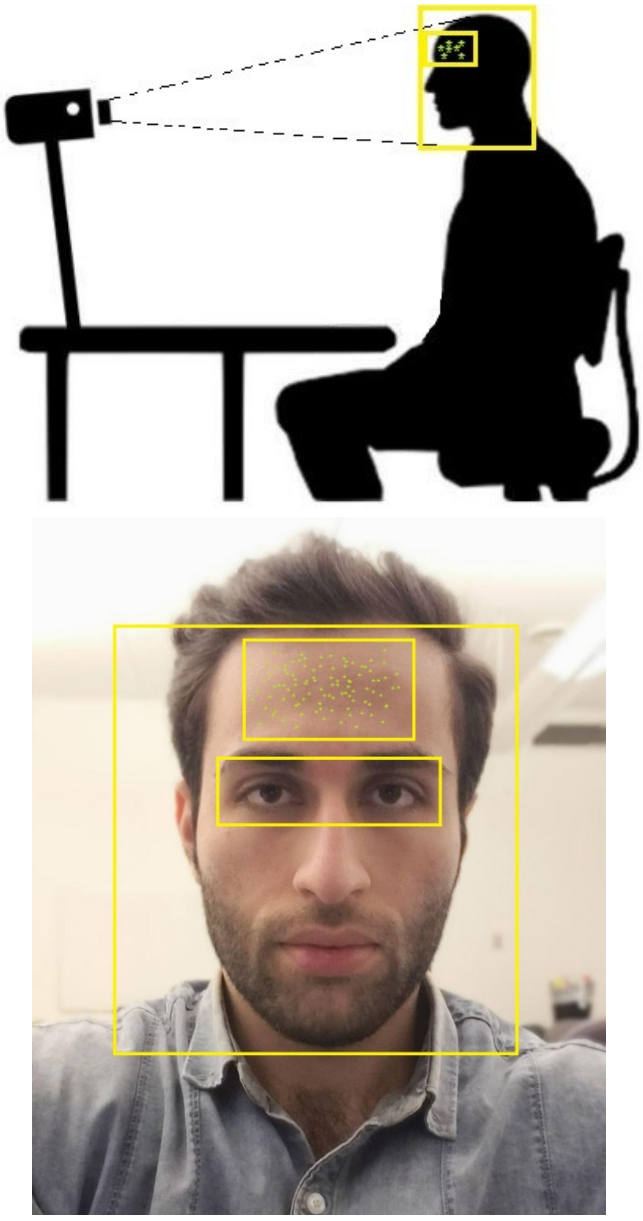


***Figure 12: Experimental Setup for Stationary Mode***

Table 3 describes the conditions for the Stationary Mode subset collection.

*Table 3: Description of the stationary data subsets*

| Parameters | Stationary Mode |
|---|---|
| **Experiment length per subject (second)** | 300 |
| **# of participants** | 10<br>(3 F ,7 M) |
| **Distance from the camera (meter)** | 1 |
| **Recording rate (fps)** | 30 |
| **# of 1 min. video segments** | 50 |
| **# of frames** | 90000 |

### 4.1.2. Movement Mode

For the movement mode, the subjects were free to roam within a 7m×7m room (walking, lying down, etc.) for a 20 minutes' video recording. We directed the subjects to face the camera regardless of their position in the room. This ensured that the subjects' forehead was visible to the webcam at all time. The lighting condition, type of camera, and ground truth calculation procedure are identical to those of the stationary mode as described in Section 4.1.1. Table 4 describes the conditions for the Movement Mode subset collection and figure 13 depicts the experimental setup for movement mode.

*Table 4: Description of the Movement data subsets*

| Parameters | Movement Mode |
|---|---|
| **Experiment length per subject (second)** | 1200 |
| **# of participants** | 10<br>(3 F ,7 M) |
| **Distance from camera (meter)** | 0-5 |
| **Recording rate (fps)** | 30 |
| **# of 1 min. video segments** | 200 |
| **# of frames** | 360000 |

*Figure 13: Experimental Setup for Movement Mode*

### 4.1.3. Dataset Training

To train the ML algorithm for IMF selection, we randomly chose 75% of the dataset (i.e. 189 one-minute-video segments) and reserved 25% to test the trained model. We processed the videos according to the procedure depicted in Fig. 9 until we reached the CEEMDAN stage. After selecting the first three IMFs in the range of interest, we extracted a total of 567 selected IMFs from 189 video segments (three IMFs from each video) of the training dataset. We labeled one of the IMFs extracted from each video as "one" and the others as "zero". The "one" label corresponds to the IMF that contains the RR information. We label the others as "zero". For accurate labeling, we calculated the absolute difference between the RR estimated from IMF and the RR obtained from the ground truth. The IMF that had the closest estimation to the ground truth is consequently labeled as "one". The other two corresponding to the same video are labeled as "zero".

Before initiating the training, in the pre-processing procedure, we shuffled the data to maintain a generalized model. Tables 5 and 6 present the cross-validation results of the ML models for each classifier and their stacked models respectively.

***Table 5: Performance of the cross-validation of individual classifiers***

| **Performance Metrics** | **Rotation Forest** | **Random Forest** | **K-Star** |
|---|---|---|---|
| **Precision** | 0.809 | 0.831 | 0.802 |
| **Recall** | 0.802 | 0.825 | 0.795 |
| **Accuracy (%)** | 80.246 | 82.469 | 79.506 |
| **Area under ROC Curve (AUC)** | 0.892 | 0.892 | 0.887 |

*Table 6: Performance of the cross-validation of stacked classifiers*

| Performance Metrics | Random Forest, Rotation Forest | K-Star, Rotation Forest | K-Star, Random Forest, Rotation Forest | Random Forest, K-Star |
|---|---|---|---|---|
| **Precision** | 0.829 | 0.817 | 0.807 | 0.845 |
| **Recall** | 0.825 | 0.810 | 0.800 | 0.842 |
| **Accuracy (%)** | 82.46 | 80.98 | 80.00 | 84.197 |
| **Area under ROC Curve (AUC)** | 0.895 | 0.875 | 0.878 | 0.904 |

## 4.2. Evaluation

With having the dataset described in stationary and movement mode, in this section, we evaluate the proposed contributions.

### 4.2.1. RR Estimation in Stationary Mode

For the stationary mode, we used the stationary subject's data subset described in section 4.1.1. In both modes, one of the three selected IMFs outputted by CEEMDAN contains the RR information. Hence, we proposed an ML model to estimate the IMF that best reflects the RR. Tables 6 and 7 show the classification results for the considered classifiers in ML models, namely Rotation Forest, Random Forest, K-Star and their stacked classifiers. Sacking classifiers improves the overall performance of the model [111]. Hence, we obtained the best performance by stacking Random Forest and K-Star. Therefore, we adopted this stacked classifier as our proposed ML-based method.

*Table 7: Performance of the individual classifiers*

| Performance Metrics | Rotation Forest | Random Forest | K-Star |
|---|---|---|---|
| **Precision** | 0.797 | 0.825 | 0.761 |
| **Recall** | 0.783 | 0.822 | 0.756 |
| **Accuracy (%)** | 78.271 | 82.222 | 75.555 |
| **Area under ROC Curve (AUC)** | 0.877 | 0.867 | 0.860 |

*Table 8: Performance of the stacked classifiers*

| Performance Metrics | Random Forest, Rotation Forest | K-Star, Rotation Forest | K-Star, Random Forest, Rotation Forest | Stacking C (Random Forest, K-Star) |
|---|---|---|---|---|
| **Precision** | 0.823 | 0.781 | 0.794 | 0.834 |
| **Recall** | 0.822 | 0.778 | 0.793 | 0.830 |
| **Accuracy (%)** | 82.22 | 77.77 | 79.25 | 82.963 |
| **Area under ROC Curve (AUC)** | 0.885 | 0.853 | 0.851 | 0.889 |

To measure rPPG, Poh et al. [57] heuristically chose the second component of the ICA as the informative channel. However, Poh et al. [6] selected the component with the highest integrated power spectrum. Sanyal and Nundy [7] used the Hue channel. Ghanadian et al. [71] proposed an ML model to choose the informative ICA channel. In Table 8, we apply CEEMDAN and the proposed ML-based method (Stacked Random Forest and K-Star) for IMF selection on the PPG channels produced by Poh et al. [57], [6], Sanyal and Nundy [7], and

Ghanadian et al. [71]. We observe from Table 8 that the combination of the Ghanadian et al. [71] rPPG collection scheme and proposed CEEMDAN with ML-based IMF selection method renders the best results. Hence, we adopt this combination as the overall proposed method for the measurement of remote RR.

***Table 9: Application of the Proposed CEEMDAN with ML-based IMF selection method on the rPPG obtained from Poh et al. [57], Poh et al. [6], Sanyal and Nundy [7] and Ghanadian et al. [71]***

| **Performance Metrics** | **Poh et al. [57] + CEEMDAN with ML IMF selection** | **Poh et al. [6] + CEEMDAN with ML IMF selection** | **Sanyal and Nundy [7] + CEEMDAN with ML IMF selection** | **Ghanadian el al. [71] + CEEMDAN with ML IMF selection** |
|---|---|---|---|---|
| **Precision** | 0.597 | 0.617 | 0.444 | 0.834 |
| **Recall** | 0.537 | 0.574 | 0.667 | 0.830 |
| **Accuracy (%)** | 53.703 | 57.446 | 66.666 | 82.963 |
| **Area under ROC Curve (AUC)** | 0.554 | 0.602 | 0.492 | 0.889 |

Table 7 compares Poh et al.'s [6,57] and Sanyal and Nundy's [7] techniques for RR estimation with the proposed method. The proposed method reduces the RMSE by 39.6% compared to Sanyal and Nundy [7], 51.8% compared to Poh et al. [57], and 44.42% compared to Poh et al. [6].

***Table 10: Comparison of Poh et al. [57], Poh et al. [6] Sanyal and Nundy [7] and proposed method in stationary mode***

| **Parameters** | **Poh et al [57]** | **Poh et al [6]** | **Sanyal and Nundy [7]** | **Proposed method** |
|---|---|---|---|---|
| **Mean Bias (BPM)** | -0.15 | -0.30 | -0.68 | -0.28 |
| **SD of Bias (BPM)** | 3.95 | 3.26 | 3.10 | 1.87 |
| **Upper limit (BPM)** | 7.60 | 6.70 | 5.40 | 3.40 |
| **Lower limit (BPM)** | -7.90 | -7.30 | -6.70 | -4.00 |
| **RMSE (BPM)** | 3.92 | 3.39 | 3.13 | 1.89 |
| **Pearson Corr. Coefficient** | 0.62* | 0.66* | 0.42* | 0.88* |

*: ($p<0.004$)

Table 8 shows the advantage of obtaining the rPPG signal to measure the RR using the method described in Ghanadian et al. [71] compared to the Poh et al. [6, 71] and Sanyal and Nundy [7] techniques. Making an assumption regarding the ICA component containing the most useful information, given that the signals are outputted in a random order, does not align with the nature of ICA [113]. Hence, we used the Ghanadian et al. [71] ML-based approach which selects the ICA component based on several of its characteristics rather than its order as proposed in [57], or only one of its properties as presented in [6].

Since respiration modulates both amplitude and frequency features of the PPG signal [58], we utilized CEEMDAN to decompose the selected ICA output to its intrinsic amplitude and frequency modulated (AM–FM) functions. In the movement mode, the subject is free to roam the environment. Therefore, the measured rPPG signal is more vulnerable to noise caused by lighting condition changes and movement [11]. CEEMDAN is an iterative and adaptive signal decomposition method that finds oscillatory features in the signal. Therefore, it allows us to differentiate effectively between undesired noise and RR information by eliminating uncorrelated IMFs.

The main challenge with the application of CEEMDAN involves selecting the intrinsic mode function that represents RR. After the initial selection of these functions in the range of interest, we proposed an ML-based algorithm for the selection of the CEEMDAN-outputted signal that best reflects the RR information. We employed a stacked ML algorithm to combine a group of "weaker learners" to build a "stronger learner". After implementing several classifiers, we achieved the best results by stacking the Random Forest and K-Star (Table. 7). Random Forest classifiers combine several decision trees using the bagging approach to improve the overall performance. K-Star is an instance-based learner that uses entropy for classification. Compared to other instance-based algorithms, K-star tends to be one of the best overall performers [110].

The application of ICA on the RGB channels results in the primary separation of noise from the source signal. Therefore, the IMFs obtained from the decomposed ICA output (after the disqualification of unlikely IMFs as described in Section 3.2) are not considerably affected by noise and hence exhibit similar dominant frequencies. Consequently, in some cases, even when the ML algorithm does not predict the best IMF (i.e. the IMF that returns the RR closest to the ground truth), the estimation is not entirely erroneous and does carry RR information. This contributes to the low RMSE we obtained in Table 9. Moreover, Poh et al. [6] calculate the RR from the HRV signal. However, some oscillations around the 0.1 Hz frequency in the HRV signal might be a result of fluctuations in the baroreceptor reflex and blood pressure [114]. Hence, the measurement of slow RR (e.g. around 6 breaths per minute) might be adversely affected by this phenomenon thus leading to erroneous RR readings [70].

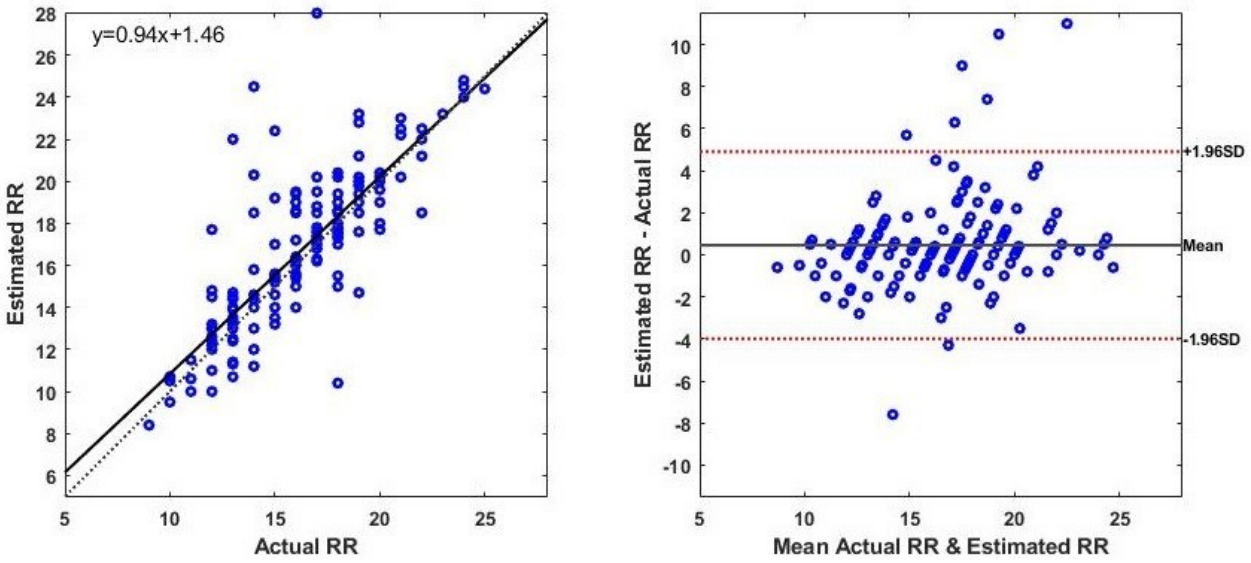
#### 4.2.2. RR Estimation in Movement Mode

To the best of our knowledge, this is the first work that proposes a remote RR measurement from RGB camera during movement. Table 10 shows the results of the proposed algorithm when we permitted the subjects to move in the room as described in Section 4.1.2. The RMSE for the proposed method in the movement mode increases by 17.8% compared to the stationary mode. However, the proposed method still exhibits a lower RMSE in the movement mode compared to those obtained for Sanyal and Nundy [7], Poh et al. [6], and Poh et al. [57] in the stationary mode. To evaluate the performance of the proposed method in movement mode, in addition to Table 10, we present scatter plots in Fig. 14. We assess the linear regression between the actual (i.e. ground truth) and estimated RR to evaluate the agreement between them. We obtained a high agreement ($R^2$=0.94) between both measures, which demonstrates that the RR measures obtained through the proposed method are highly in agreement with the ground truth.

*Table 11: Estimated RR in Movement Mode*

| Parameters | Proposed method |
|---|---|
| **Mean Bias (BPM)** | 0.46 |
| **SD of Bias (BPM)** | 2.26 |
| **Upper limit (BPM)** | 4.90 |
| **Lower limit (BPM)** | -4.01 |
| **RMSE** | 2.30 |
| **Pearson Corr. coefficient** | 0.80* |

*: ($p<0.004$)




(a) (b)

*Figure 14: Statistical Results of (a) Regression model (b) Bland and Altman*

## 4.3. Summary of Chapter 4

We recruited 10 consenting adult subjects (3 females and 7 males) to recorded 250 one-minute-videos of subjects in stationary and movement modes. The subjects wore a thoracic expansion sensor to measure RR, based on respiratory-derived chest movements. Moreover, we provided them with a thumb counter to count their RR. We adopted as ground truth the subject-measured

RR recorded by the thumb counter.

After processing the collected videos, we select the first three IMFs in the range of interest. To selecting the best IMF that reflects the RR, we utilize ML. We extract the RR from the latter IMF.

To the best of our knowledge, this is the first work that proposes a remote RR measurement from RGB camera during movement. The RMSE for the proposed method in the movement mode was 2.30 BPM which increases by 17.8% compared to the stationary mode.

# Chapter 5. Conclusion & Future Works

In this chapter, we present our conclusions and describe potential future work in this field.

## 5.1. Conclusions

We proposed an algorithm for rPPG-based RR estimation. The proposed approach uses CEEMDAN for signal decomposition. It uses an ML model for selecting the IMF outputted by CEEMDAN that best represents RR information. The ML technique selects the best IMF with an accuracy of 82.96%. Moreover, our approach exhibits more accurate results for RR estimation compared to existing state-of-the-art methods for stationary mode measurement. We demonstrated the performance of the proposed method for subjects in movement where we obtained an RSME of 2.30 BPM.

We distinguish our work from existing ones as follows:

- We apply the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) [13] to decompose the rPPG signal and obtain the most informative coefficient that reflects RR.
- We develop a Machine Learning (ML) algorithm to select the RR information carrying IMF after applying the CEEMDAN method on the selected rPPG signal.

## 5.2. Future Works

There are several potential improvements that can be made to the proposed method:

1. Deep artificial neural networks (ANNs) can be used for feature extraction from the IMF signals as opposed to engaging in feature engineering as we have done in this work. Recently, the applications of deep convolutional neural networks (CNNs) have shown its great success in various problems of computer vision same as image classification [115-117].
2. Emotional changes such as sadness, fear, happiness and anxiety can dramatically influence RR [118]. In [119], they investigated the influences of emotions in cardiorespiratory activities. They musically induced the emotions through listening to various music in different genres. At the same time, they recorded the physiological activities (e.g. RR) of individuals and finally, they were able to distinguish the emotions and the genre of listened music through these physiological signs. Masaoka et al. [120] investigated the relationship between metal stress and respiration. Using unpleasant sounds, they found out the positive correlation between unpleasant emotions caused by mental stress and RR. Also, looking at various photographs/images, which induces emotional can change RR [118]. Therefore, for future studies, combining emotion derived facial expression estimation with the RR information can be leveraged in a multimodal affect recognition system.
3. Since the ROI in the proposed method is the forehead, this region might be covered by hair or wearable artifacts. Moreover, the subject might not be facing the camera which reduces the visibility of the ROI. Therefore, the development of a robust multi-modal ROI detection scheme that can choose among multiple possible ROIs (face, cheeks, forehead etc.)

depending on the context or the application of multi-camera based system are suitable future research directions.

4. In this thesis, we do not investigate the feasibility of this technique in a real-time scenario. Therefore, deploying the proposed method in a real-time system may interest researchers in the field and a has the potential of further studies. For instance, in real-time systems it can be used for:
    - Monitoring and assessing the vital sign of athletes (e.g. soccer players) and the functionality of their lungs/respiratory system during competitions for further physiological analysis.
    - Monitoring prisoners in cells in attempt to inhibit them from suicide.
    - Monitoring the emotion/vital sign of astronauts in international space station.
    - X-ray medical imaging centers to detect the most appropriate time for exposing X-rays (with detecting the inhalation and exhalation phases) to capture the picture of chest bones.
    - Evaluating the performance of actors/actresses and their ability in believability of their role through assessing their emotion using camera.
    - Highly secured systems for assessing the emotional status of individuals (e.g. lie detector)

# Bibliography


[1] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. F. Pimentel, A. E. W. Johnson, J. Alastruey, L. Tarassenko, P. J. Watkinson, R. Beale, and D. A. Clifton, "Breathing Rate Estimation From the Electrocardiogram and Photoplethysmogram: A Review," *IEEE Reviews in Biomedical Engineering, vol*. 11, pp. 2–20, 2018.

[2] C. W. Seymour, "Prediction of Critical Illness During Out-of-Hospital Emergency Care," Jama, vol. 304, no. 7, p. 747, 2010.

[3] F. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, "Respiration rate monitoring methods: A review," *Pediatric Pulmonology*, vol. 46, no. 6, pp. 523–529, 2011.

[4] Y. Chee, J. Han, J. Youn, and K. Park, "Air mattress sensor system with balancing tube for unconstrained measurement of respiration and heart beat movements," *Physiological Measurement*, vol. 26, no. 4, pp. 413–422, 2005

[5] M. Villarroel, S. Davis, P. Watkinson, A. Guazzi, K. Mccormick, L. Tarassenko, J. Jorge, A. Shenvi, and G. Green, "Continuous non-contact vital sign monitoring in neonatal intensive care unit," Healthcare Technology Letters, vol. 1, no. 3, pp. 87–91, 2014.

[6] M.-Z. Poh, D. J. Mcduff, and R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011.

[7] S. Sanyal and K. K. Nundy, "Algorithms for Monitoring Heart Rate and Respiratory Rate From the Video of a User's Face," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1–11, 2018.

[8] R. Janssen, W. Wang, A. Moço, and G. D. Haan, "Video-based respiration monitoring with automatic region of interest detection," *Physiological Measurement*, vol. 37, no. 1, pp. 100–114, 2015.

[9] A. Heinrich, F. V. Heesch, B. Puvvula, and M. Rocque, "Video based actigraphy and breathing monitoring from the bedside table of shared beds," *Journal of Ambient Intelligence and Humanized Computing*, vol. 6, no. 1, pp. 107–10, 2014.

[10] M. H. Li, A. Yadollahi, and B. Taati, "A non-contact vision-based system for respiratory rate estimation," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.

[11] L. Feng, L.-M. Po, X. Xu, Y. Li, and R. Ma, “Motion-Resistant Remote Imaging Photoplethysmography Based on the Optical Properties of Skin,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 5, pp. 879–891, 2015.

[12] P. H. Charlton, T. Bonnici, L. Tarassenko, J. Alastruey, D. A. Clifton, R. Beale, and P. J. Watkinson, “Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants,” *Physiological Measurement*, vol. 38, no. 5, pp. 669–690, 2017.

[13] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.

[14] J. F. Fieselmann, M. S. Hendryx, C. M. Helms, and D. S. Wakefield, “Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients,” Journal of General Internal Medicine, vol. 8, no. 7, pp. 354–360, 1993.

[15] C. P. Subbe, R. G. Davies, E. Williams, P. Rutherford, and L. Gemmell, “Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions*,” Anaesthesia, vol. 58, no. 8, pp. 797–802, 2003.

[16] D. R. Goldhill, A. F. Mcnarry, G. Mandersloot, and A. Mcginley, “A physiologically-based early warning score for ward patients: the association between score and outcome,” Anaesthesia, vol. 60, no. 6, pp. 547–553, 2005.

[17] M. Cretikos, J. Chen, K. Hillman, R. Bellomo, S. Finfer, and A. Flabouris, “The objective medical emergency team activation criteria: A case–control study,” Resuscitation, vol. 73, no. 1, pp. 62–72, 2007.

[18] Shneerson, John M. Sleep medicine: a guide to sleep and its disorders. John Wiley & Sons, 2009.

[19] Y. Nishida, T. Hori, T. Suehiro, and S. Hirai, “Monitoring of breath sound under daily environment by ceiling dome microphone,” SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. Cybernetics Evolving to Systems, Humans, Organizations, and their Complex Interactions (Cat. No.00CH37166).

[20] Werthammer, Joseph, Jerome Krasner, John DiBenedetto, and Ann R. Stark. "Apnea monitoring by acoustic detection of airflow." Pediatrics 71, no. 1 (1983): 53-55.

[21] M. Folke, L. Cernerud, M. Ekström, and B. Hök, “Critical review of non-invasive respiratory monitoring in medical care,” Medical & Biological Engineering & Computing, vol. 41, no. 4, pp. 377–383, 2003.

[22] K. Storck, M. Karlsson, P. Ask, and D. Loyd, “Heat transfer evaluation of the nasal

thermistor technique," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 12, pp. 1187–1191, 1996.

[23] Folke, M., Granstedt, F., Hök, B. and Scheer, H., 2002. Comparative provocation test of respiratory monitoring methods. Journal of clinical monitoring and computing, 17(2), pp.97-103.

[24] Lee-Chiong LT, editor. Sleep: a comprehensive handbook. JohnWiley and Sons Inc.; 2006

[25] K. Nepal, E. Biegeleisen, and T. Ning, "Apnea detection and respiration rate estimation through parametric modelling," *Proceedings of the IEEE 28th Annual Northeast Bioengineering Conference (IEEE Cat. No.02CH37342)*.

[26] C. Merritt, H. Nagle, and E. Grant, "Textile-Based Capacitive Sensors for Respiration Monitoring," IEEE Sensors Journal, vol. 9, no. 1, pp. 71–78, 2009.

[27] T. Kang, "Textile-embedded sensors for wearable physiological monitoring systems," Ph.D. dissertation, North Carolina State Univ., Raleigh, NC, 2006.

[28] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiological Measurement, vol. 28, no. 3, 2007. 2007.

[29] J. Lázaro, E. Gil, R. Bailón, A. Mincholé, and P. Laguna, "Deriving respiration from photoplethysmographic pulse width," *Medical & Biological Engineering & Computing*, vol. 51, no. 1-2, pp. 233–242, 2012.

[30] P. Leonard, "Standard pulse oximeters can be used to monitor respiratory rate," *Emergency Medicine Journal*, vol. 20, no. 6, pp. 524–525, 2003.

[31] D. Wertheim, C. Olden, E. Savage, and P. Seddon, "Extracting respiratory data from pulse oximeter plethysmogram traces in newborn infants," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 94, no. 4, 2008

[32] J. Lazaro, Y. Nam, E. Gil, P. Laguna, and K. H. Chon, "Smartphone-camera-acquired pulse photoplethysmographic signal for deriving respiratory rate," *2014 8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO)*, 2014.

[33] Moody, George B., Roger G. Mark, Marjorie A. Bump, Joseph S. Weinstein, Aaron D. Berman, Joseph E. Mietus, and Ary L. Goldberger. "Clinical validation of the ECG-derived respiration (EDR) technique." *Computers in cardiology* 13 (1986): 507-510.

[34] B. Mazzanti, C. Lamberti, and J. D. Bie, "Validation of an ECG-derived respiration monitoring method," *Computers in Cardiology, 2003*, 2003.

[35] C. Caro and J. Bloice, "Contactless Apnœa Detector Based On Radar," *The Lancet*, vol. 298, no. 7731, pp. 959–961, 1971.

[36] K.-M. Chen, D. Misra, H. Wang, H.-R. Chuang, and E. Postow, "An X-Band Microwave Life-Detection System," *IEEE Transactions on Biomedical Engineering*, vol. BME-33, no. 7, pp. 697–701, 1986.

[37] C. Li, J. Ling, J. Li, and J. Lin, "Accurate Doppler Radar Noncontact Vital Sign Detection Using the RELAX Algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 3, pp. 687–695, 2010.

[38] B. Iyer, N. P. Pathak, and D. Ghosh, "RF sensor for smart home application," *International Journal of System Assurance Engineering and Management*, vol. 9, no. 1, pp. 52–57, 2016.

[39] M. Uenoyama, T. Matsui, K. Yamada, S. Suzuki, B. Takase, S. Suzuki, M. Ishihara, and M. Kawakami, "Non-contact respiratory monitoring system using a ceiling-attached microwave antenna," Medical & Biological Engineering & Computing, vol. 44, no. 9, pp. 835–840, 2006.

[40] T. Kondo, T. Uhlig, P. Pemberton, and P. D. Sly, "Laser monitoring of chest wall displacement," *European Respiratory Journal*, vol. 10, no. 8, pp. 1865–1869, 1997.

[41] S. D. Min, J. K. Kim, H. S. Shin, Y. H. Yun, C. K. Lee, and M. Lee, "Noncontact Respiration Rate Measurement System Using an Ultrasonic Proximity Sensor," *IEEE Sensors Journal*, vol. 10, no. 11, pp. 1732–1739, 2010.

[42] A. P. Prathosh, P. Praveena, L. K. Mestha, and S. Bharadwaj, "Estimation of Respiratory Pattern From Video Using Selective Ensemble Aggregation," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2902–2916, 2017.

[43] M. Yang, Q. Liu, T. Turner, and Y. Wu, "Vital sign estimation from passive thermal video," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[44] R. Murthy, I. Pavlidis, and P. Tsiamyrtzis, "Touchless monitoring of breathing function," *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.

[45] C. B. Pereira, X. Yu, V. Blazek, and S. Leonhardt, "Robust remote monitoring of breathing function by using infrared thermography," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.

[46] U. Wijenayake and S.-Y. Park, "Real-Time External Respiratory Motion Measuring Technique Using an RGB-D Camera and Principal Component Analysis," *Sensors*, vol. 17, no. 8, p. 1840, 2017.

[47] K.-Y. Lin, D.-Y. Chen, and W.-J. Tsai, "Image-Based Motion-Tolerant Remote Respiratory Rate Evaluation," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 3263–3271, 2016.

[48] Q.-V. Tran, S.-F. Su, and V.-T. Nguyen, "Pyramidal Lucas-Kanade-Based Noncontact Breath Motion Detection," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*,

pp. 1–12, 2018.

[49] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, “Remote Measurements of Heart and Respiration Rates for Telemedicine,” *PLoS ONE*, vol. 8, no. 10, 2013.

[50] D. Shao, Y. Yang, C. Liu, F. Tsow, H. Yu, and N. Tao, “Noncontact Monitoring Breathing Pattern, Exhalation Flow Rate and Pulse Transit Time,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2760–2767, 2014.

[51] A. Al-Naji and J. Chahl, “Remote respiratory monitoring system based on developing motion magnification technique,” *Biomedical Signal Processing and Control*, vol. 29, pp. 1–10, 2016.

[52] B. Wei, X. He, C. Zhang, and X. Wu, “Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions,” *BioMedical Engineering OnLine*, vol. 16, no. 1, 2017.

[53] J. Jorge, M. Viillarroel, S. Chaichulee, K. Mccormick, and L. Tarassenko, “Data fusion for improved camera-based detection of respiration in neonates,” *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, 2018.

[54] X. Yang and T. Bourlai, “Video-Based Human Respiratory Wavelet Extraction and Identity Recognition,” *Surveillance in Action Advanced Sciences and Technologies for Security Applications*, pp. 51–75, 2017.

[55] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, “Eulerian video magnification for revealing subtle changes in the world,” ACM Transactions on Graphics, vol. 31, no. 4, pp. 1–8, 2012.

[56] K. Alghoul, S. Alharthi, H. A. Osman, and A. E. Saddik, “Heart Rate Variability Extraction From Videos Signals: ICA vs. EVM Comparison,” IEEE Access, vol. 5, pp. 4711–4719, 2017.

[57] M.-Z. Poh, D. J. Mcduff, and R. W. Picard, “Non-contact, automated cardiac pulse measurements using video imaging and blind source separation,” *Optics Express*, vol. 18, no. 10, p. 10762, 2010.

[58] W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino, and G. A. Dumont, “Estimation of Respiratory Rate From Photoplethysmographic Imaging Videos Compared to Pulse Oximetry,” *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1331–1338, 2015.

[59] P. Comon, “Independent component analysis, A new concept?” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[60] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light,” Opt. Exp., vol. 16, no. 26, pp. 21434–21445, 2008.

[61] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter Respiratory Rate Estimation From the Photoplethysmogram," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–1953, 2013.

[62] M. V. Gastel, S. Stuijk, and G. D. Haan, "Robust respiration detection from remote photoplethysmography," *Biomedical Optics Express*, vol. 7, no. 12, p. 4941, 2016.

[63] M. A. Motin, C. K. Karmakar, and M. Palaniswami, "Ensemble Empirical Mode Decomposition With Principal Component Analysis: A Novel Approach for Extracting Respiratory Rate and Heart Rate From Photoplethysmographic Signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 766–774, 2018.

[64] 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 auto-regressive models," *Physiological Measurement*, vol. 35, no. 5, pp. 807–831, 2014.

[65] D. F. Dietrich, C. Schindler, J. Schwartz, J.-C. Barthélémy, J.-M. Tschopp, F. Roche, A. V. Eckardstein, O. Brändli, P. Leuenberger, D. R. Gold, J.-M. Gaspoz, and U. Ackermann-Liebrich, "Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the SAPALDIA study," *EP Europace*, vol. 8, no. 7, pp. 521–529, 2006.

[66] M. R. Everingham, J. Sivic, and A. Zisserman, "Hello! My name is... Buffy -- Automatic Naming of Characters in TV Video," *Procedings of the British Machine Vision Conference 2006*, 2006

[67] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*.

[68] OpenCV Library. Accessed: Sep. 18, 2016. [Online]. Available: http://opencv.org/

[69] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[70] D. P. Giddens and R. I. Kitney, "Neonatal heart rate variability and its relation to respiration," Journal of Theoretical Biology, vol. 113, no. 4, pp. 759–780, 1985.

[71] H. Ghanadian, M. Ghodratigohar, and H. A. Osman, "A Machine Learning Method to Improve Non-Contact Heart Rate Monitoring Using an RGB Camera," *IEEE Access*, vol. 6, pp. 57085–57094, 2018.

[72] E. Hjelmås and B. K. Low, "Face Detection: A Survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236–274, 2001

[73] Tolba, A. S., A. H. El-Baz, and A. A. El-Harby. "Face recognition: A literature review." International Journal of Signal Processing 2, no. 2 (2006): 88-103.

[74] Brimblecombe, Phil. "Face detection using neural networks." H615–Meng Electronic Engineering, School of Electronics and Physical Sciences, URN 1046063 (2002).

[75] Hjelmås, Erik, and Boon Kee Low. "Face detection: A survey." Computer vision and image understanding 83, no. 3 (2001): 236-274.

[76] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," *Proceedings CVPR 89: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[77] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.

[78] Z. Jin, Z. Lou, J. Yang, and Q. Sun, "Face detection using template matching and skin-color information," Neurocomputing, vol. 70, no. 4-6, pp. 794–800, 2007.

[79] Z. Wang and S. Li, "Face Recognition using Skin Color Segmentation and Template Matching Algorithms," Information Technology Journal, vol. 10, no. 12, pp. 2308–2314, 2011.

[80] Yang, Ming-Hsuan, Narendra Ahuja, and David Kriegman. "A survey on face detection methods." (1999).

[81] S. J. Mckenna, Y. Raja, and S. Gong, "Object tracking using adaptive colour mixture models," *Computer Vision — ACCV98 Lecture Notes in Computer Science*, pp. 615–622, 1997.

[82] M.-H. Yang and N. Ahuja, "Detecting human faces in color images," *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*.

[83] Hunke, H. Martin. Locating and tracking of human faces with neural networks. Carnegie-Mellon University. Department of Computer Science, 1994.

[84] H. Azami, K. Mohammadi, and B. Bozorgtabar, "An Improved Signal Segmentation Using Moving Average and Savitzky-Golay Filter," *Journal of Signal and Information Processing*, vol. 03, no. 01, pp. 39–44, 2012.

[85] Lee, Han-Wook, Ju-Won Lee, Won-Geun Jung, and Gun-Ki Lee. "The periodic moving average filter for removing motion artifacts from PPG signals." International Journal of Control, Automation, and Systems 5, no. 6 (2007): 701-706.

[86] D. Mcduff, S. Gontarek, and R. W. Picard, "Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera," *IEEE Transactions on Biomedical*

*Engineering*, vol. 61, no. 10, pp. 2593–2601, 2014.

[87] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical Optics Express*, vol. 6, no. 5, p. 1565, 2015

[88] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[89] Q. Wang, P. Yang, and Y. Zhang, "Artifact reduction based on Empirical Mode Decomposition (EMD) in photoplethysmography for pulse rate detection," *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010.

[90] K. V. Madhav, M. R. Ram, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "Estimation of respiration rate from ECG, BP and PPG signals using empirical mode decomposition," *2011 IEEE International Instrumentation and Measurement Technology Conference*, 2011.

[91] Garde, A., Walter Karlen, P. Dehkordi, J. M. Ansermino, and G. A. Dumont. "Empirical mode decomposition for respiratory and heart rate estimation from the photoplethysmogram." In Computing in Cardiology Conference (CinC), 2013, pp. 799-802. IEEE, 2013.

[92] Z. Wu and N. E. Huang, "Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method," *Advances in Adaptive Data Analysis*, vol. 01, no. 01, pp. 1–41, 2009.

[93] M. A. Motin, C. K. Karmakar, and M. Palaniswami, "Ensemble Empirical Mode Decomposition With Principal Component Analysis: A Novel Approach for Extracting Respiratory Rate and Heart Rate From Photoplethysmographic Signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 766–774, 2018.

[94] C. Orphanidou, "Derivation of respiration rate from ambulatory ECG and PPG using Ensemble Empirical Mode Decomposition: Comparison and fusion," Computers in Biology and Medicine, vol. 81, pp. 45–54, 2017.

[95] M. R. and S. Prabhu, "A Novel Algorithm to Obtain Respiratory Rate from the PPG Signal," *International Journal of Computer Applications*, vol. 126, no. 15, pp. 9–12, 2015.

[96] Michalski, Ryszard S., Jaime G. Carbonell, and Tom M. Mitchell, eds. Machine learning: An artificial intelligence approach. Springer Science & Business Media, 2013.

[97] A. Talwar and Y. Kumar, "Machine Learning: An artificial intelligence methodology," *Int. J. Eng. Comput. Sci.*, vol. 2, no. 12, pp. 3400–3404, 2013.

[98] Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. "Reinforcement learning: A survey." Journal of artificial intelligence research 4 (1996): 237-285.

[99] S. Fallet, V. Moser, F. Braun, and J.-M. Vesin, "Imaging Photoplethysmography: What are the Best Locations on the Face to Estimate Heart Rate?," *2016 Computing in Cardiology Conference (CinC)*, 2016.

[100] M. Tarvainen, P. Ranta-Aho, and P. Karjalainen, "An advanced detrending method with application to HRV analysis," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, pp. 172–175, 2002.

[101] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical Mode Decomposition as a Filter Bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.

[102] M. A. Colominas, G. Schlotthauer, M. E. Torres, and P. Flandrin, "Noise-Assisted Emd Methods In Action," *Advances in Adaptive Data Analysis*, vol. 04, no. 04, p. 1250025, 2012.

[103] S. Fleming, M. Thompson, R. Stevens, C. Heneghan, A. Plüddemann, I. Maconochie, L. Tarassenko, and D. Mant, "Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies," *The Lancet*, vol. 377, no. 9770, pp. 1011–1018, 2011.

[104] F. Song, Z. Guo, and D. Mei, "Feature Selection Using Principal Component Analysis," *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, 2010.

[105] Meyer, Yves. Wavelets and operators. Vol. 1. Cambridge university press, 1992.

[106] Box, George EP, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

[107] Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

[108] Horning, Ned. "Random Forests: An algorithm for image classification and generation of continuous fields data sets." In Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan, vol. 911. 2010.

[109] J. Rodriguez, L. Kuncheva, and C. Alonso, "Rotation Forest: A New Classifier Ensemble Method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1619–1630, 2006.

[110] J. G. Cleary and L. E. Trigg, "K*: An Instance-based Learner Using an Entropic Distance Measure," Machine Learning Proceedings 1995, pp. 108–114, 1995.

[111] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: from early developments to recent advancements," *Systems Science & Control Engineering*, vol. 2, no. 1, pp. 602–609,

2014.

[112] Vital Signs (Body Temperature, Pulse Rate, Respiration Rate, Blood Pressure), Johns Hopkins Medicine Health Library. Accessed: Apr. 28, 2018.[Online].Available:https://www.hopkinsmedicine.org/healthlibrary/conditions/adult/cardiovascular_diseases/vital_signs_body_temperature_pulse_rate_respiration_rate_blood_pressure_85,P00866

[113] J. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 6, p. 362, 1993.

[114] R. W. Deboer, J. M. Karemaker, and J. Strackee, "Comparing Spectra of a Series of Point Events Particularly for Heart Rate Variability Data," IEEE Transactions on Biomedical Engineering, vol. BME-31, no. 4, pp. 384–387, 1984.

[115] Luo W, Li Y, Urtasun R, Zemel R. Understanding the effective receptive field in deep convolutional neural networks. InAdvances in neural information processing systems, pp. 4898-4906, 2017.

[116] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[117] O'Shea K, Nash R. "An introduction to convolutional neural networks". arXiv preprint arXiv:1511.08458. 2015 Nov 26.

[118] I. Homma and Y. Masaoka, "Breathing rhythms and emotions," Experimental Physiology, vol. 93, no. 9, pp. 1011–1021, 2008.

[119] Nyklíček I, Thayer JF, Van Doornen LJ. "Cardiorespiratory differentiation of musically-induced emotions. Journal of Psychophysiology". 1997.

[120] Y. Masaoka and I. Homma, "Anxiety and respiratory patterns: their relationship during mental stress and physical load," International Journal of Psychophysiology, vol. 27, no. 2, pp. 153–159, 1997.