Final Master's Thesis
MASTER'S DEGREE
IN
BIOMEDICAL ENGINEERING

Assessment of the accuracy and reliability of heart rate and blood oxygen saturation measurement in wearable devices

Barcelona June 12th, 2023

Author: Antoni López Giménez
Director: Juan José Ramos Castro

Carried out at: Biomedical Electronics and Instrumentation Group (IEB) - ETSETB (UPC)

UNIVERSITAT DE BARCELONA

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Registration Date: 28/4/23

MASTER IN BIOMEDICAL ENGINEERING

Final Master's Project Record

STUDENT NAME: Antoni López Giménez

TITLE OF THE PROJECT:

Assessment of the accuracy and reliability of heart rate and blood oxygen saturation measurement in wearable devices

DIRECTOR AND TUTOR OF THE PROJECT:
(Specify Center, Department, Company, Hospital, etc. where they are associated)

Director/Tutor: Juan José Ramos Castro, ETSETB-UPC. Department: CREB-IEB

Signed:

JUAN JOSE RAMOS CASTRO - DNI 43411651W

Firmado digitalmente por JUAN JOSE RAMOS CASTRO - DNI 43411651W
Nombre de reconocimiento (DN): c=ES, sn=RAMOS CASTRO, givenName=JUAN JOSE, serialNumber=IDCES-43411651W, cn=JUAN JOSE RAMOS CASTRO - DNI 43411651W
Fecha: 2023.05.05 17:33:06 +02'00'

Dr/a.
Project's Director & tutor

Signed:

Student

Signed:

Digitally signed by JORDI FONOLLOSA MAGRINYA - DNI ***5987** (TCAT), on 25/05/2023, with a digital certificate issued by EC-SectorPublic

Dr. Jordi Fonollosa
UPC Coordinator
MSc in Biomedical Engineeing

Signed:

CASTAÑO LINARES OSCAR - 52408576V

Dr. Óscar Castaño
UB Coordinator
MSc in Biomedical Engineering

ABSTRACT

Wearable devices that can measure heart rate (HR) and blood oxygen saturation (SpO2) non-invasively and continuously are becoming increasingly popular for both fitness and medical purposes. However, measuring these vital signs accurately and reliably in wearable devices is challenging due to various factors, such as device design, sensor technology, signal processing, user behaviour, and environmental conditions. This thesis aims to assess the accuracy and reliability of HR and SpO2 measurement in the wearable device LifeVit Vital, which uses photoplethysmography (PPG) sensors to measure blood volume changes in the wrist. The thesis compares the results of LifeVit Vital with gold-standard methods such as electrocardiography (ECG) and oximetry, trough different positions (lie down, sit down, stand up, sport). The thesis uses the Bland-Altman method to analyse the agreement between the methods and discusses the challenges and limitations of using PPG-based sensors for HR and SpO2 measurement. The main findings of this thesis were that LifeVit Vital shows good agreement with ECG for HR measurement at rest, but not during physical activity, and that LifeVit Vital shows poor agreement with oximetry for SpO2 measurement in all situations. In this will be suggested some possible improvements for LifeVit Vital and future directions for research in this field.

Index

Figures Index

Tables Index

Glossary

BLE	Bluetooth Low Energy
BPM	Beats Per Minute
CE	Conformité Européenne (European Conformity)
ECG	Electrocardiography
EMC	Electromagnetic Compatibility
EU	European Union
HR	Heart Rate
LED	Light-Emitting Diode
LoA	Limits of Agreement
NaN	Not a Number
PPG	Photoplethysmography
QRS	QRS Complex
SDK	Software Development Kit
SpO2	Blood Oxygen Saturation

1. INTRODUCTION

The greatest problem for new companies or start-ups in biomedical engineering is the budget and the money that they can spend on new devices. Another aspect to consider is that any device, as wearables, or software with the aim to be used on medical environments must pass the medical device certification and it must have the CE marking as it is explained in Regulation EU 2017/745 [1].

If you look at the wearable markets, in Europe, all of them pass the CE marking but most of them not as medical device, in other countries the CE marking is not mandatory. That may sound contradictory, but it is not. The CE marking is almost mandatory in all electronic devices, independently of the aim, whether there is a directive or regulation of free movement in Europe. The aim of the CE is free circulation within the European community. In conclusion, it is a mark that represents pursuant to the requirements of a European directive or regulation, which may be safety and operational requirements.

On the other hand, the medical device certification is a specific regulation for all these electronic devices that based on the Regulation EU 2017/745 medical device definition that says [1]: *"'medical device' means any instrument, apparatus, appliance, software, implant, reagent, material, or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes:*

- *diagnosis, prevention, monitoring, prediction, prognosis, treatment, or alleviation of disease,*
- *diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability,*
- *investigation, replacement, or modification of the anatomy or of a physiological or pathological process or state*
- *providing information by means of in vitro examination of specimens derived from the human body, including organ, blood, and tissue donations,*

and which does not achieve its principal intended action by pharmacological, immunological, or metabolic means, in or on the human body, but which may be assisted in its function by such means.

The following products shall also be deemed to be medical devices:

- *devices for the control or support of conception.*
- *products specifically intended for the cleaning, disinfection or sterilisation of devices as referred to in Article 1(4) and of those referred to in the first paragraph of this point."*

That certification corroborates that the devices is safe, accurate and reliable. This certification is expensive, strict and takes a long time, for that reason the great number of manufacturers take the decision to not use with medical aim or sell them to non-medical uses.

1.1. Statement of purpose

For all the previous information, the startup DyCare needed an assessment of their new acquisition, LifeVit Vital smartwatch, to have information of the accuracy in the heart rate and blood oxygen saturation measurements and have some confidence with the objective to implement LifeVit Vital as an accessory for their medical device software, ReHub (DyCare telerehabilitation platform).

In other words, we planned a qualitative assessment to see if the results obtained by the device can be accurate and similar to results obtained with expensive devices or methods as ECG with Biopac System. That assessment is the consequence of LifeVit company passed CE marking according to an EMC or electrical safety directive but is not CE marked as a medical device and DyCare did not have the confidence of this device measurement.

1.2. State of the art

Heart rate (HR) and blood oxygen saturation (SpO2) are two vital signs that reflect the cardiovascular and respiratory health of a person. HR is the number of times the heart beats per minute, and SpO2 is the percentage of oxygenated haemoglobin in the blood. Monitoring these parameters can help detect and prevent various diseases, such as arrhythmias, heart failure, hypoxia, and COVID-19.

Wearable devices that can measure HR and SpO2 non-invasively and continuously are becoming increasingly popular for both fitness and medical purposes. These devices can provide real-time feedback, personalized recommendations, and remote monitoring for users and clinicians, potentially improving health outcomes and quality of life [2].

However, measuring HR and SpO2 accurately and reliably in wearable devices is challenging due to several factors, such as device design, sensor technology, signal processing, user behaviour, and environmental conditions [3].

Most wearable devices use optical sensors based on photoplethysmography (PPG) to measure HR and SpO2. PPG is a technique that uses light-emitting diodes (LEDs) to illuminate the skin and photodetectors to measure the changes in light absorption caused by blood volume pulsations. PPG signals can be obtained from different body locations, such as the finger, earlobe, wrist, or forehead. However, wrist-worn devices are more convenient and comfortable for users than other locations. PPG signals can be classified into two types: transmissive and reflective [4].

Transmissive PPG uses LEDs and photodetectors on opposite sides of the tissue, such as the finger or earlobe, to measure the light transmitted through the tissue. Reflective PPG uses LEDs and photodetectors on the same side of the tissue, such as the wrist or forehead, to measure the light reflected from the tissue. Reflective PPG is more suitable for wearable devices than transmissive PPG because it does not require a tight contact between the sensor and the skin.

To measure HR from PPG signals, wearable devices use algorithms to detect the peaks or valleys of the pulsatile component of the signal, which corresponds to the cardiac cycle. The time intervals between consecutive peaks or valleys are then used to calculate the HR. To measure SpO2 from PPG signals, wearable devices use algorithms to estimate the ratio of oxygenated to deoxygenated haemoglobin in the blood based on the different absorption characteristics of red and infrared light by haemoglobin. The ratio is then converted to SpO2 using a calibration curve.

However, PPG signals are prone to noise and artifacts caused by motion, ambient light, skin pigmentation, temperature, blood pressure, and other factors These factors can affect the quality and reliability of HR and SpO2 measurements in wearable devices. Therefore, various methods have been proposed to improve the accuracy and robustness of PPG-based HR and SpO2 measurements in wearable devices. One method is to use multiple wavelengths of light to obtain more information about the blood volume changes and haemoglobin absorption in different tissue layers. For example, green light can penetrate deeper into the skin than red or infrared light, and thus can capture more blood volume changes in the arteries than in the veins or capillaries. [3]

Some wearable devices use green LEDs along with red or infrared LEDs to measure HR and SpO2 more accurately. Another method is to use multiple sensors to obtain complementary or redundant information about HR and SpO2. For example, some wearable devices use electrocardiography (ECG) sensors along with PPG sensors to measure HR more accurately [2].

ECG is a technique that measures the electrical activity of the heart using electrodes attached to the skin. ECG signals can provide more reliable information, is considered the gold-standard method, about HR than PPG signals because they are less affected by motion artifacts. However, ECG sensors require a good contact between the electrodes and the skin, which may not be feasible or comfortable for some users or situations. A third method is to use advanced signal processing techniques to enhance or filter PPG signals before extracting HR or SpO2 information. For example, some techniques use adaptive filters, wavelet transforms, machine learning algorithms, or deep neural networks to reduce noise or artifacts from PPG signals [3].

A recent study assessed the agreement between two photoplethysmography-based wearable devices (Apple Watch S6 and Polar Vantage M2) for monitoring HR during different physical activity situations (lying, sitting, standing, and walking) with a gold-standard ECG system. The study involved 18 university students who wore the devices simultaneously while performing the activities for consecutive 5-min periods [5].

The study found that the Apple Watch S6 had a higher agreement with the ECG system in terms of HR measurement across all activities and averaging times, compared to the Polar Vantage M2. The study also found that both devices had lower agreement with the ECG system when the activity involved some level of physical activity, especially for shorter averaging times.

The study suggested that photoplethysmography-based wearable devices are suitable for monitoring HR averages at regular intervals, especially at rest, but their feasibility is questionable for a continuous analysis of HR for research or clinical purposes, especially when involving some level of physical activity. Also proposed a new methodology to synchronize and measure the agreement of HR measurements from different devices without interpolation or resampling.

1.3. Methodology preview

To be able to achieve the aim of this thesis and help DyCare with their purpose to improve their telerehabilitation platform we planned to do an assessment comparing the LifeVit device against gold-standard methods in both measurements. All the methodology will be explained in section 2.1 with more details.

2. PROJECT DEVELOPMENT

In this section, I will explain how it was decided to afront the assessment of the device in terms of theory and the methodology followed. We will pass through the theory behind the technology used by devices (both the device we are assessing and the reference oximeter), ECG cardiac signal representation, the method used to analyse the results (Bland-Altman Method) which is the most used in biomedical field in assessment of new medical devices and to conclude the code done to analyse the raw data and obtain the final figures.

2.1. Methodology

In this section, an attempt will be made to explain by steps all the methodology used to achieve the purpose of the thesis. It will be explained from the protocol used to contact with the volunteers to all the code and the process used to obtain the final data set with which all the results were obtained. All the annexes commented will be found in the thesis GitHub repository.

2.1.1. Bland-Altman Method

In September 1983, D.G. Altman and J.M Bland published a journal article explaining through the title "Measurement in Medicine: The Analysis of Method Comparison Studies" their methodology to compare two different methods of measuring some quantity [6].

During the paper, both statisticians were doing a review of the methods used so far like *"comparison of means", correlation* and *regression* and why were wrong methods to compare two different measurement methods or technologies. The aim of both were to find a pragmatic approach to show and explain the results to a non-statistician. At the end, the method is like a linear regression graphics taking the reference method in abscises axis and in ordinate axis the new method. One of the main limitations of the Bland-Altman method is when there is no linear relationship between the two methods. It could happen that the two methods measure with very little variability but one of them is non-linear. In this case the linearity correction would be necessary. With a linear regression we will have a graphic as shown figure 1 extracted by the original document:

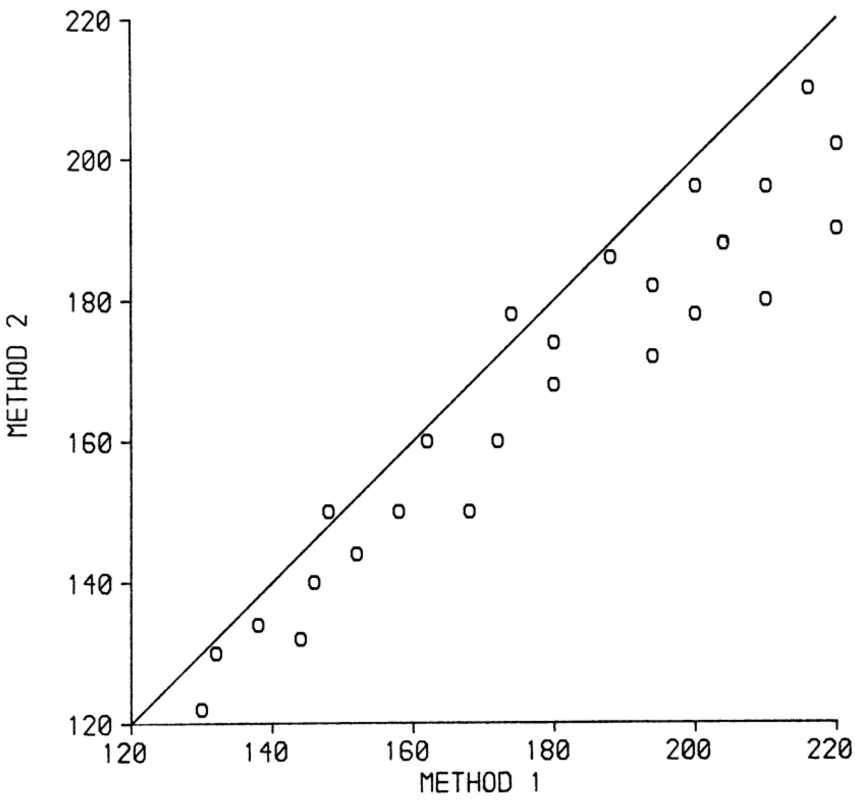

Figure 1 Comparison of two methods with linear regression

After that, they present their specific method that represents in abscises axis the average of two methods and in the ordinate axis the difference between methods. As in the previous method I extracted the example from the original document:

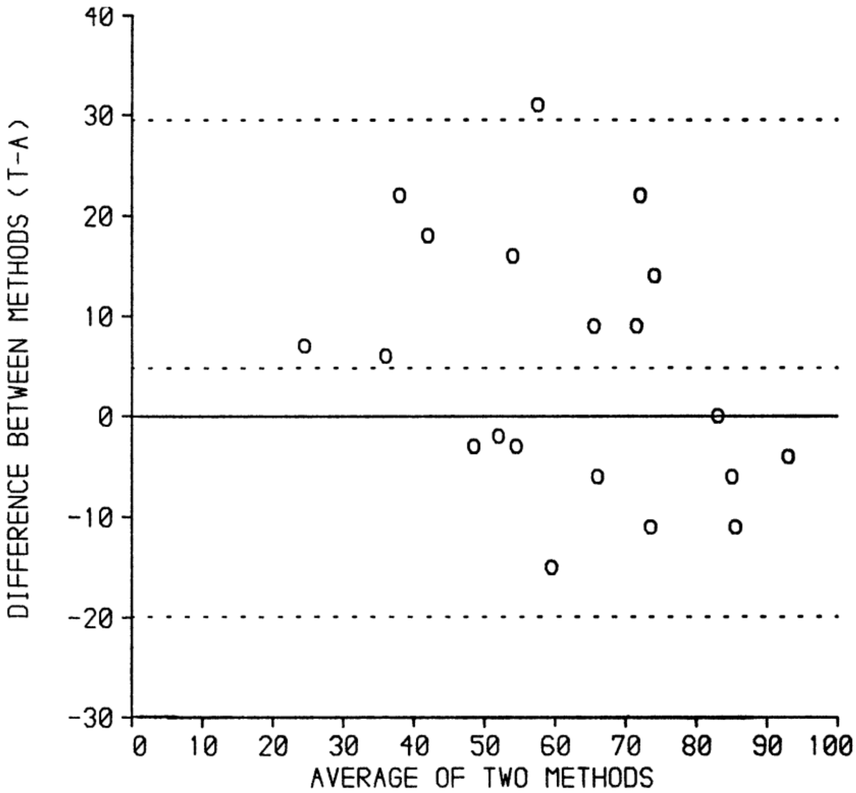

Figure 2 Comparison of two methods with Bland-Altman plot.

As you can appreciate in the figure 2, the results of the two-measurement methods difference can be seen clearly. The function that can describe the Bland-Altman graphic could be:

$$F(x, y) = \left(\frac{M2 + M1}{2}, M2 - M1\right)$$

In the previous equation, M1 is the reference method and M2 the new method to assess. Also, it must be explained the new dotted lines that appears in figure 2. The middle line is the mean that in other terms is the statistical/estimated bias between the methods. The outer dotted lines represent the 95% of confidence for the experiment. To obtain them it must be multiplied by 1.96 the standard deviation of the experiment.

However, the Bland-Altman method relies on assumptions such as normally distributed paired differences, constant bias, and variance homogeneity. When these assumptions are not met, it must be suggested use non-parametric quantile estimators for estimating the 2.5% and 97.5% quantiles of the differences. One recent study compares several non-parametric quantile estimators and finds that some perform well for small sample sizes while others perform well for larger sample sizes. Consequently, the article provides evidence for the use of non-parametric methods for estimating confidence intervals when errors are not normally distributed [7].

2.1.2. Protocol

The document where the protocol is explained with more detail can be found in annex A of this thesis. The protocol indicates the steps followed by the researcher to obtain all the measurements and only suffers a little change. A minor adjustment to the protocol was done, since most of our volunteers are also working with a PhD student, the protocol was harmonized with hers.

The change was that: the volunteers performed all the experiment blindfolded and hearing white noise to isolate them from the environment excluding sport experiment of this isolation. Furthermore, we increase the number of positions to 4 adding lie down position which results in: lie down, sit down, stand up and sport.

The protocol consisted of doing measurements simultaneously with the wearable to assess, LifeVit Vital, ECG measured by Biopac System and MP36 hardware and finally the heart rate and SpO2 with the Wellue oximeter. Every experiment takes 10 min except for sports one that only takes 6 min. After every experiment step/position the results were saved in Matlab extension to analyse then with it. The only different extension was the wearable one, the measurements were saved in a .txt file in a smartphone with the Android SDK developed by LifeVit.

The population size for the assess was agreed with DyCare and the advisor of this thesis in 15-20 volunteers. At the end were found 22 volunteers that they were part of the experiment. Initially, all the people independently of the sex, age range or geographic-ethnic were good for the experiment but during the experiments was found that some people with very dark skin can increase bad results. Nor were there any sensitive groups or anyone who might be affected by such technology.

2.1.3. Code

Therefore, that the greatest part of the code developed has a few hundred lines it is complicated to add images or directly the code in this thesis. All the code will be delivered with the aim to be checked all the work done. Even though, the code is properly commented a short introduction and explanation will be done in this section. All the code can be found in the thesis GitHub repository.

2.1.3.1. Android SDK

The Android SDK was full developed by LifeVit company with the aim to give to their customers a base code to know how implement some functionalities and how the device works. The original SDK can be found in a GitHub repository and the modified class for this thesis could be found in the SW folder.

Only was modified the specified class of the device of interest which can found in the with the following name BraceletVitalActivity.java. The changes done were:

- Reduction of the options to manage only leaving those of interest.
- Creation of new variables and function to save the measurements.
- Creation of schedules to be able to measure 3 min each parameter. This change was done because the original SDK and the device are limited to only measure for one minute.

2.1.3.2. Spo2_Wellue_reference.m

This Script was created to be able to obtain all the measurements done by the Wellue oximeter. This code was done based on the information found in the following GitHub repository https://github.com/roger-/FS20F/blob/main/parser.py. This previous script has a lot of information of how parse the bytes send by the device and the BLE characteristic to read it. Finally, all the information parsed was added to a timetable.

2.1.3.3. Analysis_Script.m

This code consists of four main sections: pre-analysis, analysis, post-analysis and functions. In the pre-analysis section, the code loads the data from different sources, such as .mat and .txt files, and applies some filtering and peak detection techniques to obtain the heart rate and respiration signals from the ECG and Biopac data. The code also creates synthetic timestamps for the data based on the LifeVit Vital device's reference time.

In the analysis section, the code adjusts the signals for possible delays and aligns them with a common evaluation time window. The code also removes any zero-peaks or NaN values that may affect the analysis. Then, the code interpolates the signals to have the same number of values for each device and plots them to check the quality of the interpolation. Finally, the code creates new tables with the final heart rate and SpO2 data for each device and position of the volunteer.

In the post-analysis section, the code saves the new tables to .mat files for further processing. In the functions section, the code defines two custom functions: one that transforms the LifeVit Vital .txt file to MATLAB format (transformLifevitFile ()), and another that detects the QRS complexes (qrsDetector(y,fs)) from the ECG signal using a Butterworth filter and a thresholding method. The function qrsDetector() is an adaptation of the classic Pan-Thompkins algorithm [8].

2.1.3.4. BA_Method_Script.m

This script performs the Bland-Altman method. This code is used to compare the two measurements (HR and SpO2) of the final data and assess their agreement. The code has four main parts:

- Data loading: This part loads the data from two files, FinalHRData.mat and FinalSpO2Data.mat, which contain heart rate and oxygen saturation measurements from different devices. The code assigns each column of the data matrix to a variable, such as Wellue, LifeVit_Vital, Biopac, etc.

- Selecting part: This part allows the user to select or delete the part of the dataset that is not of interest. It is commented by default, but it can be used to filter out some positions or volunteers from the data.
- Previous calculations: This part calculates the differences and averages between the measurements from different devices, as well as the mean and the 95% confidence limits of the differences. These values are needed for plotting the Bland-Altman method.
- Plotting: In this part only was called the function BlanAltmanPlot() a few time to obtains all the methods of interest and the correlation values of each comparison.
- Function: blandAltmanPlot(average,difference,mean,lim,titleText) plot the differences against the averages for each pair of devices. The function also plots the mean, the mean trend and the confidence limits of the differences, and adds labels, titles and legends to the plots.

2.1.4. Data Collection and Processing

At this point, the methodology used to analyse all the raw data will be explained with the aim to obtain the desired results. As volunteer #0 had some of the best results in terms of clear data, their data was chosen with the aim of explaining the steps for obtaining the final data more clearly. This data was included in the final analysis using the Bland-Altman method.

The experiment duration was explained in previous sections and any kind of calibration was needed. The only aspect to consider was that the device must be well-adjusted as will be explained in the conclusion section.

The only difference between the volunteer #0 and the others is that the heart rate obtained with Wellue Oximeter does not appear but for our goal now, it is not necessary. The figure 3 shows all signals with their time stamps as they were recorded. As you can observe we have different delays between all the signals, this is because the time stamp recorded by the LifeVit wearable had differences (a few seconds) with the computer clock. The Biopac and Biopac 2 signals were obtained with the QRS algorithm.

Hereafter filtered with two different filters and subsequently a synthetic time stamp was added, obtained with the initiation of the LifeVit one and with the duration of the original ECG signal. Both Biopac signals were filtered with the same parameters (a windows of 10 samples) but with different Matlab functions:

1. Biopac: It was filtered with "filter()" Matlab function that allow us to filter with a given window in a only one way, from the left to the right. This process introduces a delay in the output signal of 5 samples [9].

2. Biopac 2: It was filtered with "filtfilt()" Matlab function that allow us to implement a bi-directional moving average filter that eliminates the initial transient and the signal delay of 5 samples [10].

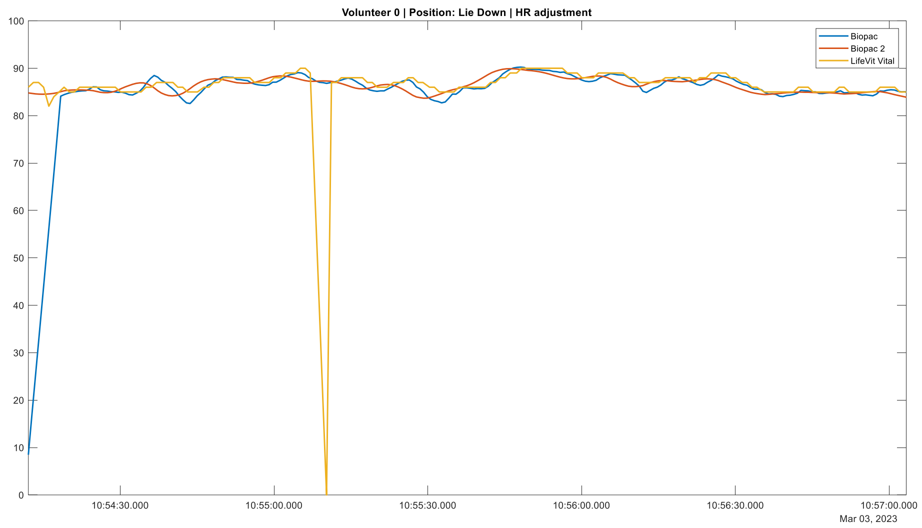

Figure 3 Raw Data filtered.

Once I analysed how much was the delay between the different signals, I decided to adjust this delay to have as much as possible synchronized signals. The result of this adjustment can be seen in the figure below. This delay depends mainly in how the heartrate signal is processed inside the Lifer Vital device. At the beginning of the acquisition, it requires several seconds to start giving heartrate values.

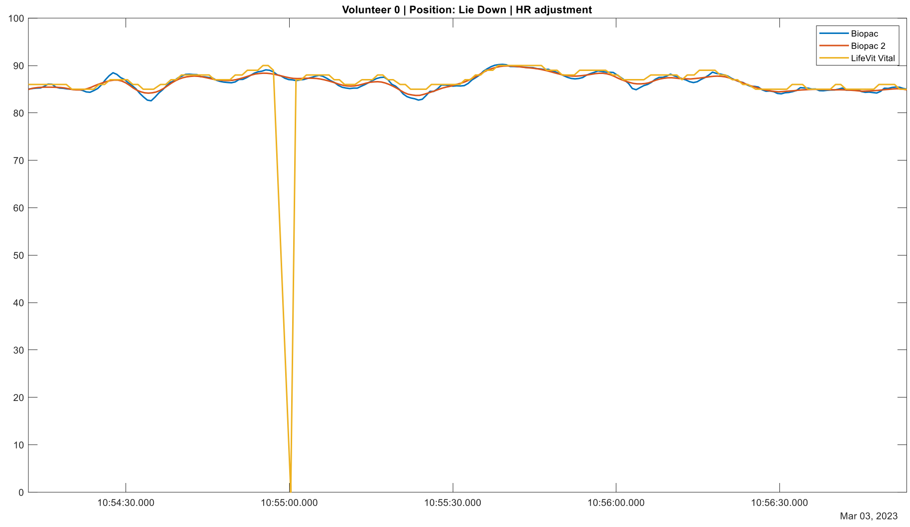

Figure 4 Data filtered once the delay has been corrected.

As you may see in figure 4, we have some "0 peaks". That is due to the original Android SDK and hardware being limited to only one-minute measurements. With the aim to assess properly the wearable, we forced the SDK to schedule every fifty-seven seconds the specific operation that we want (heart rate measurement or SpO2 measurement). Therefore, as these non-desirable peaks are not real, I removed them with the code, deleting the values at the problematic points in all the signals and I obtained the following graphic.

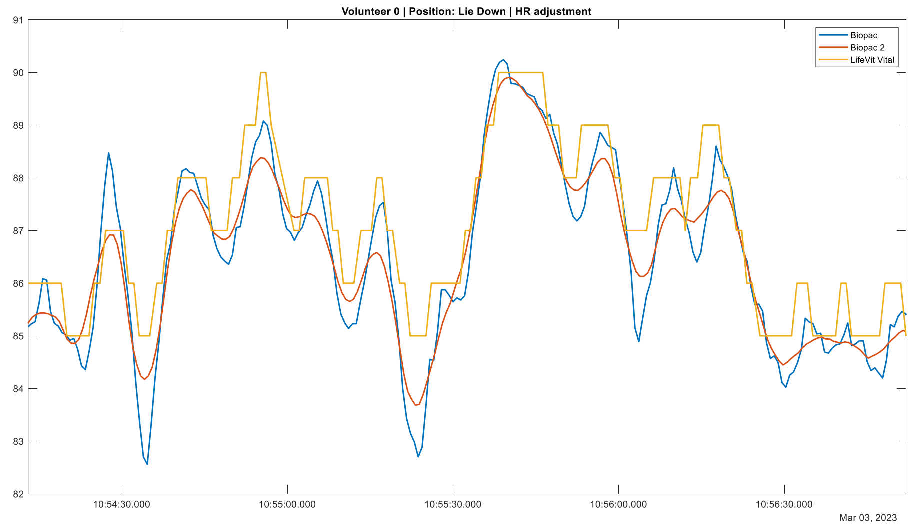

Figure 5 Final Data

As can be seen in the figure 5, the similarity between all the curves is very large. The last step to be able to compare all the signals we resampled all of them to have the same number of samples as it shown in the figure 6 below.

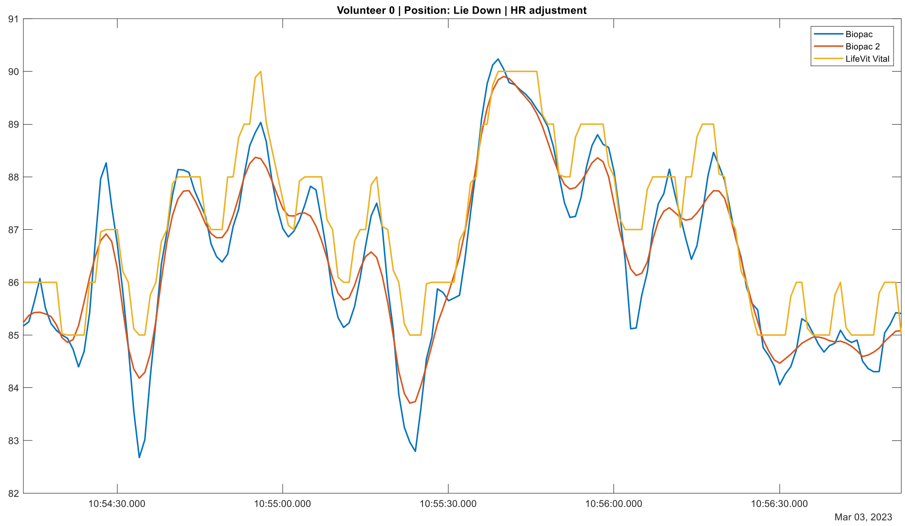

Figure 6 Final Data Resampled.

When all this analysis was done, I created a .mat file with all the results considered that can appear some "NaN" values and they are not desired, all of them were deleted previously. This "NaN" values appear due to the interpolation done to obtain the same number of samples in all the signal. Again, they were deleted together with the values in the other signals at the same position to maintain the same length.

3. RESULTS

Once all the theoretical and methodological parts have been explained it is time to show you the results obtained. To start with, we contacted volunteers by sending them the text explained in the annex A, which represents the validation protocol followed during the experiments. As you can imagine, in order not to extend this thesis too much, you can find the whole protocol explained in the annex specified before. We recollect some statistical data, with the questionnaire in annex B, from the volunteers that you can see in the table below:

Patient Number	Biological Sex	Height (cm)	Weight (kg)	BMI	Smoker	Times/Week	Kind Sport	Respirathory problems	Heart Conditions	Other Previous Pathology	Positions order
0	F	168	66	23,38	TRUE	2	Gym workout	FALSE	FALSE	FALSE	S/ST/L/W
1	M	186	79	22,84	FALSE	?	Cricket, Gym	FALSE	FALSE	FALSE	S/L/ST/W
2	M	175	82	26,78	FALSE	5	Gym	FALSE	FALSE	FALSE	ST/L/S/W
3	M	186	65	18,79	FALSE	3	Run, Gym	FALSE	FALSE	FALSE	ST/S/L/W
4	F	176	53	17,11	FALSE	1	Gym	FALSE	FALSE	FALSE	L/S/ST/W
5	M	177	70	22,34	FALSE	4	Gym,Run	FALSE	FALSE	FALSE	ST/L/S/W
6	M	170	65	22,49	FALSE	0	-	FALSE	FALSE	FALSE	S/ST/L/W
7	M	180	73	22,53	FALSE	4	Soccer, Run	FALSE	FALSE	FALSE	L/S/ST/W
8	M	178	70	22,09	FALSE	3	Gym, Run	FALSE	FALSE	FALSE	S/ST/L/W
9	F	176	70	22,60	FALSE	3	Basketball	FALSE	TRUE	FALSE	ST/L/S/W
10	F	170	85	29,41	FALSE	5	Walk	FALSE	FALSE	TRUE	S/L/ST/W
11	M	173	75	25,06	FALSE	2	Bicycling	FALSE	FALSE	FALSE	S/ST/L/W
12	M	170	57	19,72	FALSE	2	Fittness	FALSE	FALSE	FALSE	ST/S/L/W
13	M	175	78	25,47	FALSE	6	Fittness	FALSE	FALSE	FALSE	ST/S/L/W
14	M	180	75	23,15	FALSE	6	Run, Gym, Swim	FALSE	FALSE	FALSE	S/ST/L/W
15	M	180	80	24,69	TRUE	0	-	FALSE	FALSE	FALSE	S/ST/L/W
16	M	173	80	26,73	FALSE	3	Padel	FALSE	FALSE	TRUE	S/W/L
17	F	177	72	22,98	TRUE	4	Gym (Cardio)	FALSE	FALSE	FALSE	S/ST/L/W
18	F	176	68	21,95	FALSE	3	Gym (Cardio)	FALSE	FALSE	FALSE	S/ST/L/W
19	M	184	78	23,04	FALSE	5	Ski,Run	FALSE	FALSE	FALSE	ST/L/S/W
20	M	190	75	20,78	FALSE	4	Basketball	FALSE	FALSE	FALSE	S/L/ST/W
21	M	181	80	24,42	FALSE	4	Football,Gym	FALSE	FALSE	FALSE	ST/L/S/W

Table 1 List of Volunteers Statistical Data

From all the experiments what we obtained was raw data from ECG and signals with different time stamp. For that reason, we developed the code explained in section 2.1.3 with the aim to analyse all the data, create a time stamp for heart rate obtained through filtered ECG to create the final dataset with the data of interest. Also, all the methodology to obtain this final data was explained in section 2.1.4.

In the table below, you can see the data used, marked with "x", the data not used for different reasons (hardware problems, poor data collection because of hair, etc.) marked with "-" and the data not available marked with "N/A". The poor data was due to the electrodes used that did not stick well to the skin in some volunteers with hairy chest.

The reason for what the volunteers #0 and #1 didn't have data available for SpO2, as I explained before, is due to the Wellue Oximeter had not yet arrived. In the case of volunteer #16 we didn't have data in stand-up position because he is a spinal cord injured person, and it was not possible to do it.

Finally, all the squares in orange it is because they were analysed with the second improved version of the algorithm for QRS complex which help us to obtain better results in sport and with volunteer #11 too.

Volunteer	Lie Down		Sit Down		Stand up		Sport	
#	HR	SpO2	HR	SpO2	HR	SpO2	HR	SpO2
0	x	N/A	x	N/A	x	N/A	x	N/A
1	-	N/A	-	N/A	-	N/A	-	N/A
2	x	-	x	x	x	x	x	x
3	x	x	x	-	-	x	-	N/A
4	x	x	x	x	-	x	x	x
5	x	x	-	x	-	x	x	x
6	x	x	x	x	x	x	x	x
7	-	x	-	x	x	x	x	x
8	-	x	-	x	-	x	x	x
9	x	x	x	x	x	-	x	x
10	x	x	x	x	x	x	x	x
11	x	x	x	x	x	x	x	x
12	x	x	x	x	x	x	x	x
13	x	x	x	x	-	x	x	x
14	x	x	x	x	x	x	x	x
15	x	x	x	x	x	x	x	x
16	x	x	x	x	N/A	N/A	-	x
17	x	x	x	x	x	x	x	x
18	x	x	x	x	x	x	x	x
19	-	x	x	x	x	x	x	x
20	x	x	x	x	x	x	x	x
21	x	x	x	x	x	x	x	x

x	Done
-	Not Used
N/A	Not Available

Table 2 Final Data Used

3.1. Heart Rate Results

Therefore, I consider it appropriate begin with a global vision of the results previously to show you how is the wearable behaviour in every position that in my opinion can be more interesting than the global one to see how its performance is.

As you will see in the following figures the results are not so good because they include all the sport situation results. That is a problem for the results since the wearable as we will see doesn't provides good readings when the volunteer is in motion. This phenomenon appears with both measurements, heart rate and blood oxygen saturation, during the sports situation.

3.1.1. All Data

The figure 7 and figure 8 shows the heart rate results between LifeVit Vital measurements against the Biopac results filtered as it was explained in section 2.1.4. It can be appreciated that if we calculate the mean's trend does not match with the overall mean calculated, due to the sports situations results and in addition because of is the result of the linear regression.

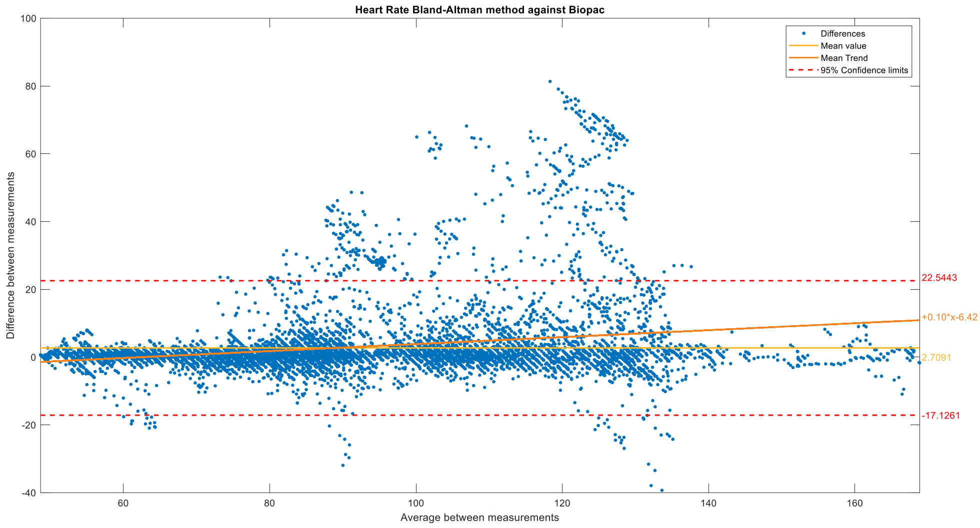

Figure 7 HR LifeVit measurements against Biopac

This slope in the mean's trend, that in the end is the result of the linear regression, can be interpreted as a gain error, due to an ideal mean line must be close to 0 or orthogonal with a short offset value. With these results can be seen that considering sports situations we will have a gain error and an increment on the standard deviation of the experiment. As can be observed this appears in both Biopac results with a similar slope (~0.10).

Furthermore, it must be careful with this gain error, as it may not reflect the true shape of the error distribution. The adjustment used to estimate the trend relies on parametric methods that may not be appropriate in this case. As can be seen in figure 12 and figure 13, removing the sport measurements, these results are improved.

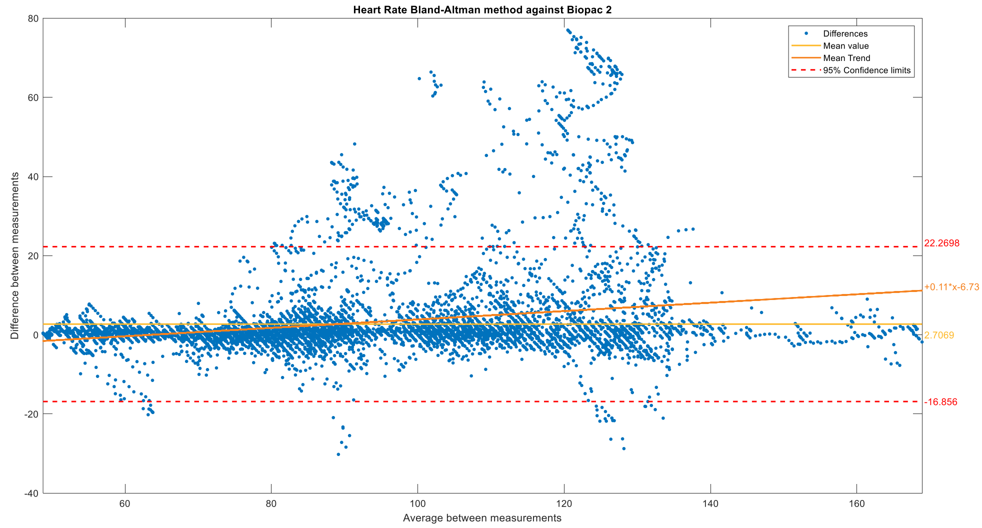

Figure 8 HR LifeVit measurements against Biopac 2

Another comparison that can be interesting is with the measurements of heart rate of the oximeter that was taken as a reference for SpO2 measurements (Figure 9). The first conclusion that can be seen is that the mean's trend slope is a little bit higher, around ~0.17 against ~0.10 in Biopac, but in addition there are more outliers. Consequently, the standard deviation of the experiment has been increased by 8 in the upper limit.

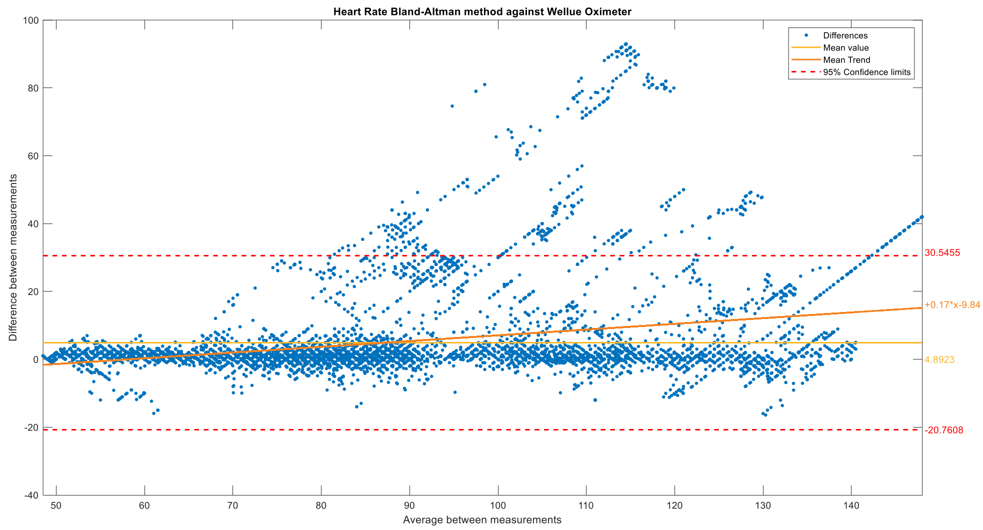

Figure 9 HR LifeVit measurements against Wellue oximeter

To be able to say that it really was the LifeVit device the problem in the heart rate measurements, considering sports situations, the method was done between Biopac results and Wellue oximeter.

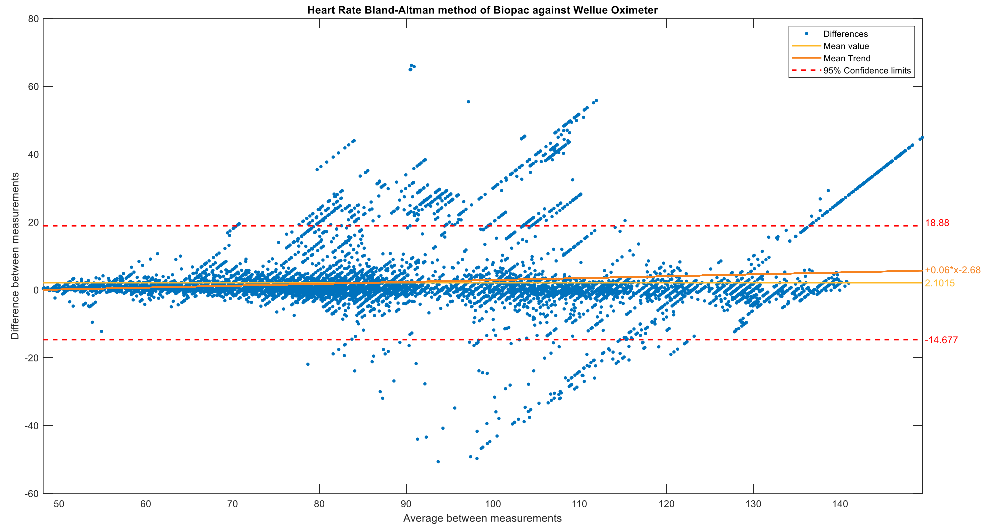

Figure 10 HR Biopac measurements against Wellue oximeter

As can be observed in figure 10 and figure 11, the improvement in the results was considerable. The overall mean value was reduced by factor 2, the mean trend slope was reduced in a 64.7%. Finally, a small difference between confidence interval limits was obtained.

This means that the problem is not on the method, or the technology used by the devices, is due to the LifeVit Vital devices have problems with heart rate measurements in motion.

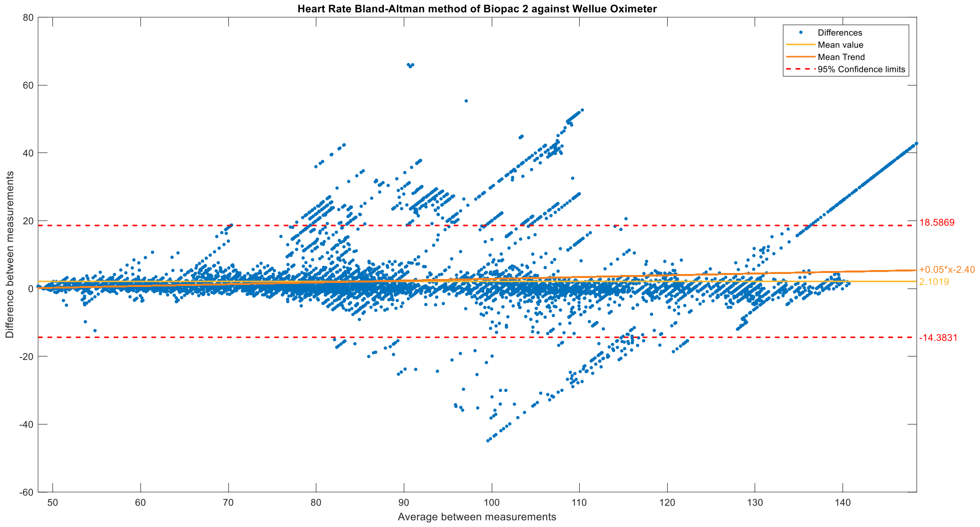

Figure 11 HR Biopac 2 measurements against Wellue oximeter

After seeing all these results and how the sports affect to the SpO2 assessment, increasing the upper and lower limit and consequently the overall mean, I decided to see the results without considering the sports measurements. In that case, when the data was analysed without the previous consideration can be seen in the figure 12 that the slope of the mean's trend was reduced in a 10 factor, from ~0.10 to ~0.01, matching more accurately with the overall mean calculated having also reduced its value.

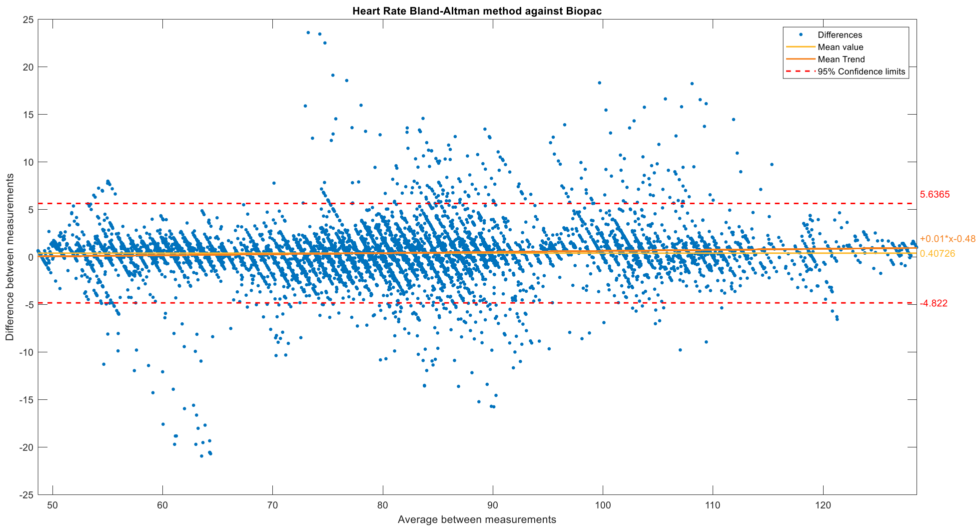

Figure 12 HR LifeVit measurements against Biopac without sport

Another effect of removing the sport recording is that the confidence interval limits were reduced significantly, and this confirm us that the wearable measures are better than the previous results. In other terms, the LifeVit Vital with heart rate measurements provides similar results as the gold standard, ECG, with a 95% confidence interval o +/-5bpm.

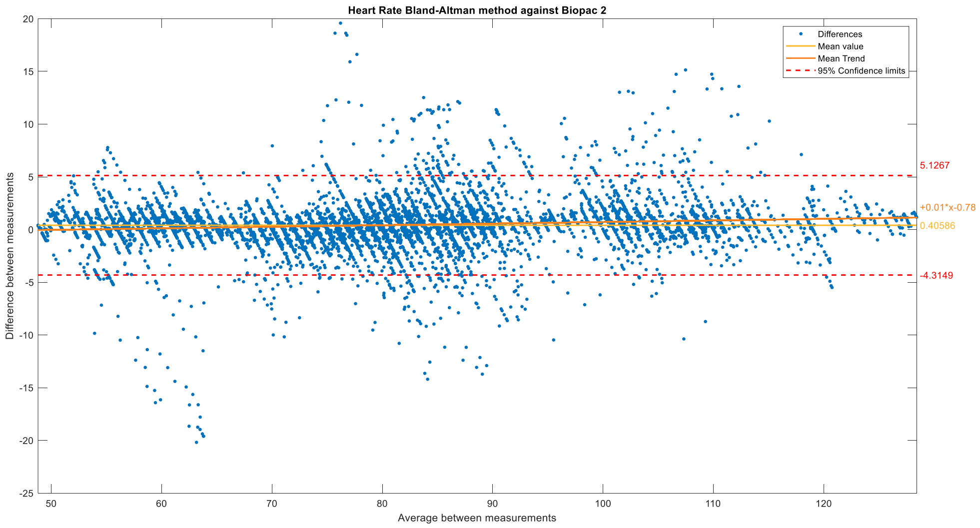

Figure 13 LifeVit measurements against Biopac 2 without sport

In the LifeVit against Wellue oximeter when comparing HR is where it can be observed more differences. This is a consequence of both devices are bad measuring during sport situations because both use the same technology to measure, photoplethysmography. Consequently, it can be observed again that the mean's trend slope and overall mean together with interval confidence limits were reduced significantly.

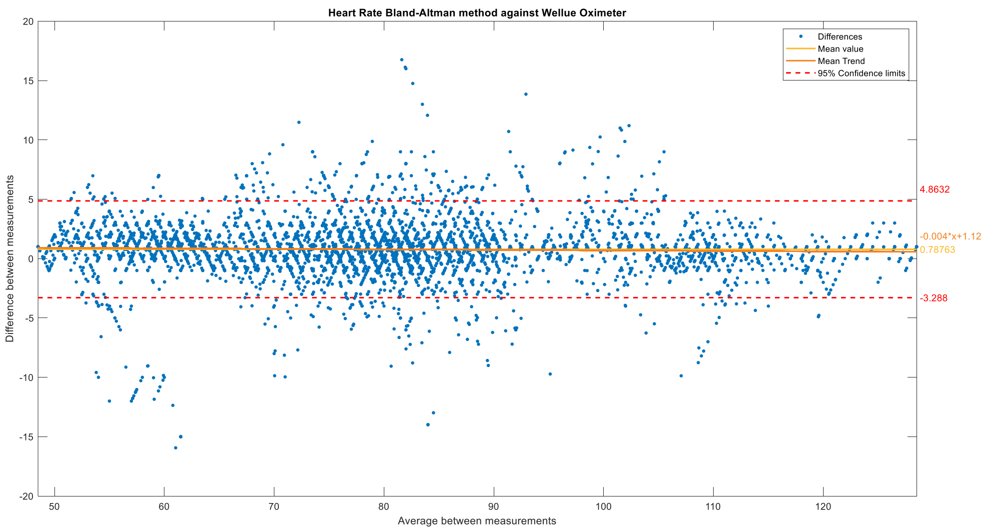

Figure 14 HR LifeVit measurements against Wellue oximeter without Sport

Once, the general results were analysed it was time to see every experiment situation independently which give us the most important information to determine how the wearable qualitatively is. The four position that will be analysed are: Lie Down, Sit Down, Stand Up and Sport.

3.1.2. Lie Down position

Firstly, in lay down position it can be observed similar results to the previous one. The main aspect observed is, the overall mean and the slope of mean's trend is the same than in previous conclusions. The only difference is that the confidence interval limits were reduced significantly because of the different distributions between the different positions.

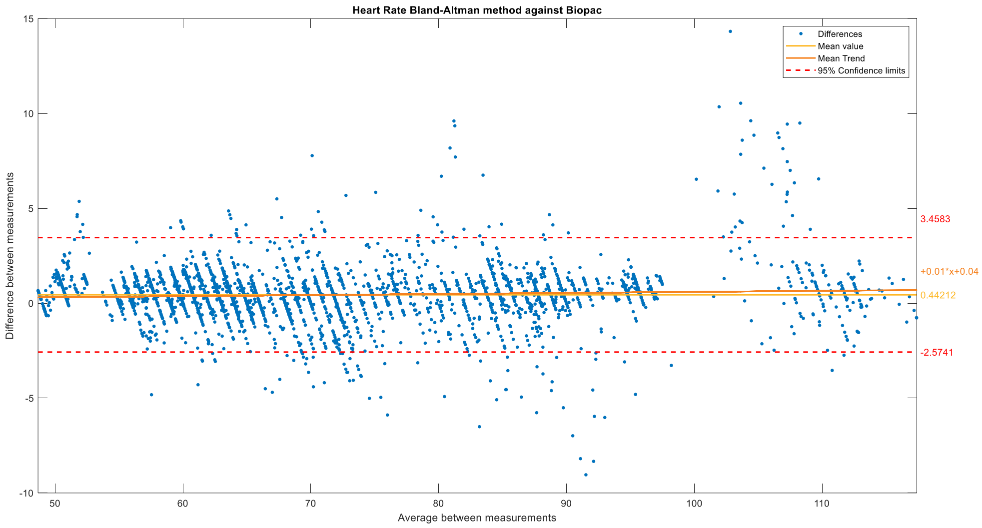

Figure 15 HR LifeVit measurements against Biopac (Lie Down Position)

In the figure 15 and 16 can be observed the conclusions commented in previous paragraph. It should be emphasised that a reduction in confidence interval limits is good news for the purpose of this thesis and the wearable.

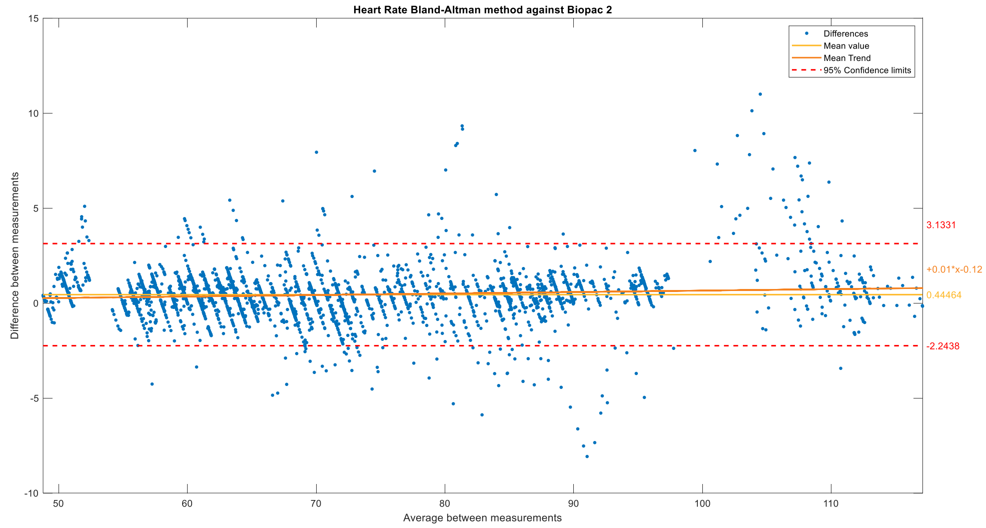

Figure 16 HR LifeVit measurements against Biopac 2 (Lie Down Position)

In the comparison of the heart rate measurements, between the LifeVit Vital and Wellue oximeter (figure 17) can be observed, the same as in Biopac results, only a small increase in the overall mean. Hereafter, the lie down results will be taken as a reference to assess the future results of the different positions.

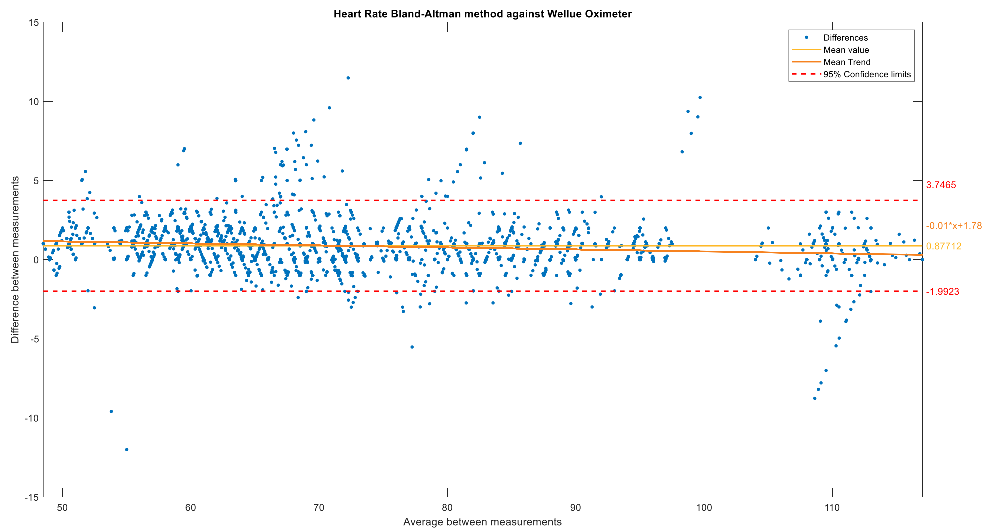

Figure 17 HR LifeVit measurements against Wellue oximeter (Lie Down Position)

3.1.3. Sit Down position

Firstly, sit down position will be analysed. The results against Biopac were good in terms of mean but with a small increase in the confidence interval. This can be observed graphically because we can see a great number of samples around 0 that reduces the overall mean but with more outliers than before which increase the standard deviation and consequently, the confidence interval limits.

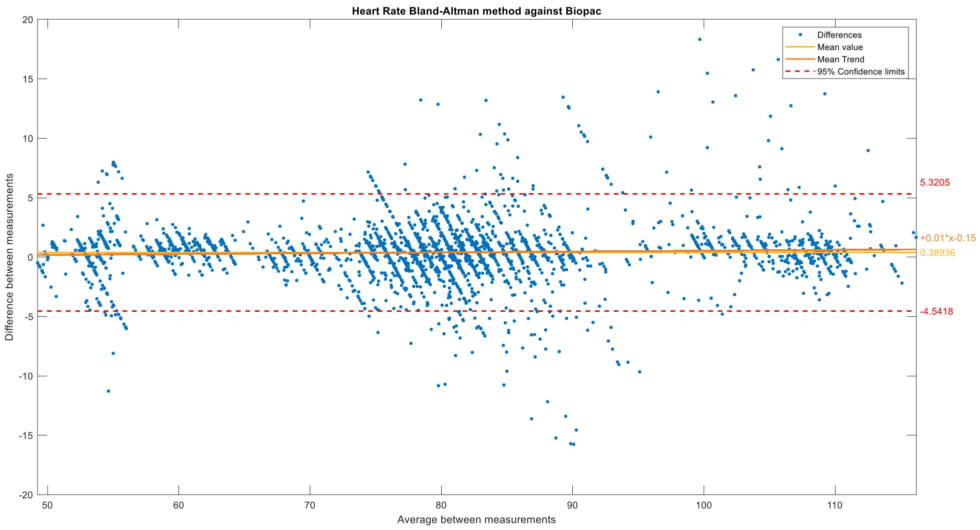

Figure 18 HR LifeVit measurements against Biopac (Sit Down Position)

In Biopac figures (figure 18 and 19) can be seen that the slope of the mean's trend is the same than lie down results but with a significant change in the intercept. This small change in the intercept could be attributed to statistical variations due the sample size.

These results must be considered in the future implementation because some patients will be monitored during breaks between exercises. In addition, in terms of heart rate is the confidence interval limit is small enough to detect bigger changes in HR to where the patient could be at risk.

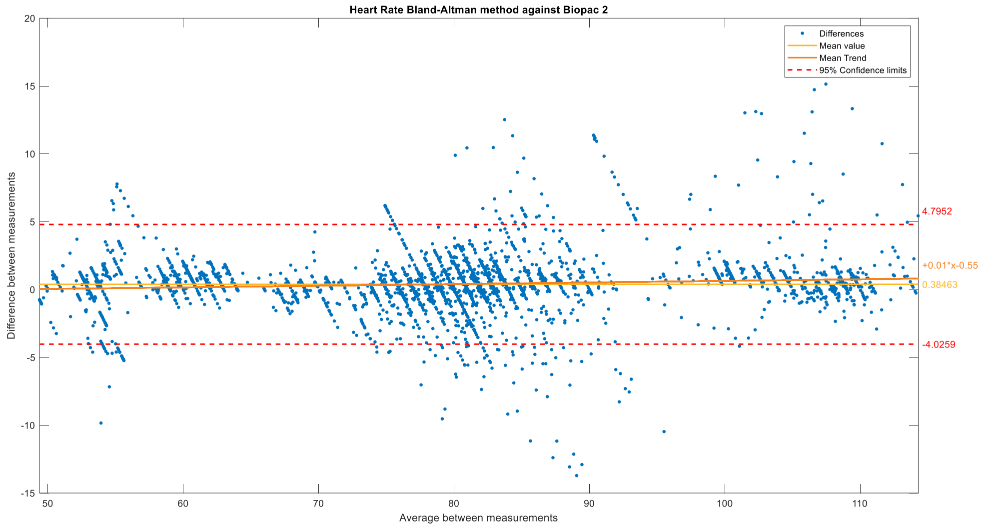

Figure 19 HR LifeVit measurements against Biopac 2 (Sit Down Position)

As in Biopac results for this position in the following figure it can be seen the HR measurements against the Wellue oximeter. They are very similar to the Biopac.

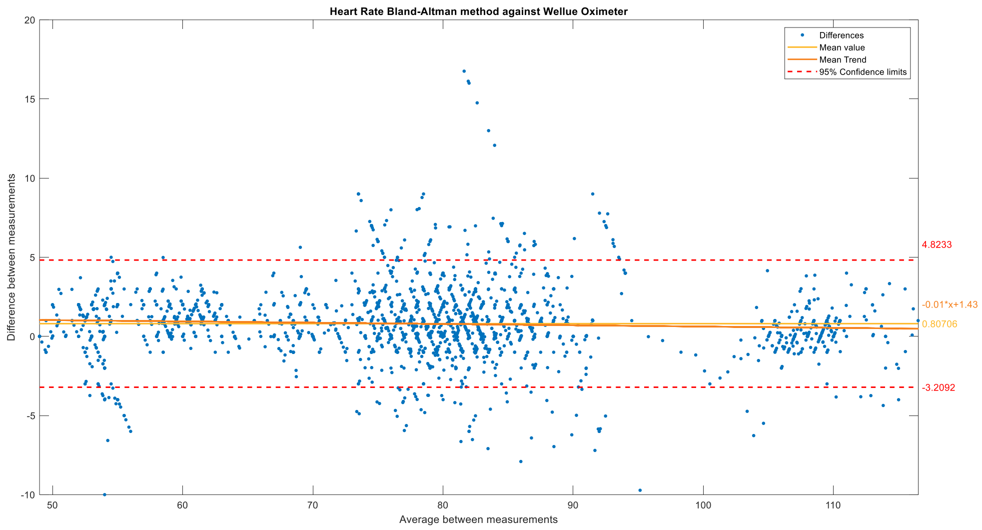

Figure 20 HR LifeVit measurements against Wellue oximeter (Sit Down Position)

3.1.4. Stand Up position

Following, the stand-up position is the other relevant situation thinking in the aim of the wearable for DyCare. If you think in rehabilitation situations is the most common posture that we can have during it.

The stand-up position can be considered as a rest position in rehabilitation. For this reason, the following results also must be compared against sit down results and not only considering lie down results. Taking into account all these previous details will be reviewed as much detailed as possible.

In the following figures we can see similar results as in Sit Down position, mean error close to zero but with an increase in the confidence interval limits. That again is due to some outliers in the instantaneous heart rate that affect directly to the standard deviation.

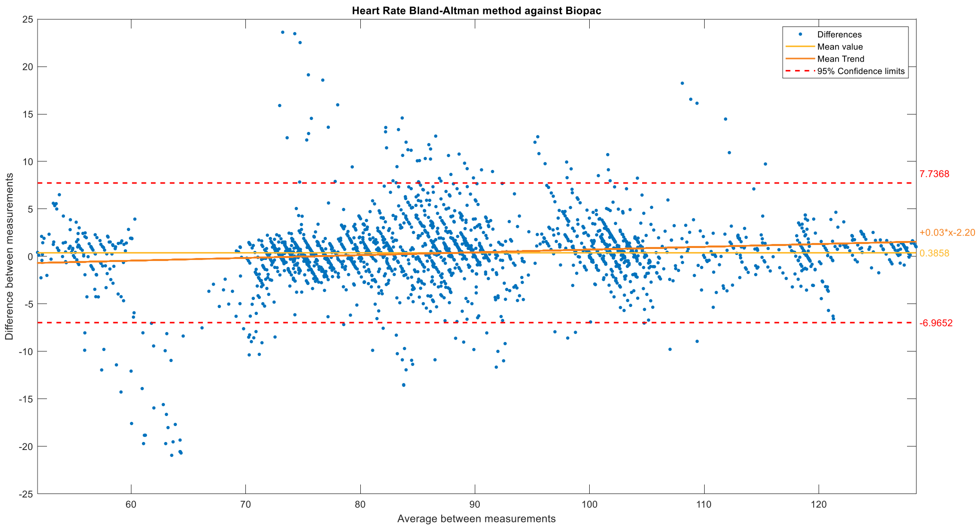

Figure 21 HR LifeVit measurements against Biopac (Stand Up Position)

Another aspect that can be seen is the mean trend slope that was multiplied by 3 in both reference cases. However, this change can be attributed again to statistical fluctuations due to the sample size. Recent studies that assess Apple Watch measurements found that a good agreement was with the following values: a standard deviation ~2.93 with the lower limit in -5.84 and upper limit 5.63 [12].

To sum up, seeing the previous results an increment in the outlier samples can be for different situations: a bad adjustment of the smartwatch, movements during the experiments, etc. These movements during the experiment were usual due to the isolation effects on the volunteers inducing destabilisation that can affect to the results.

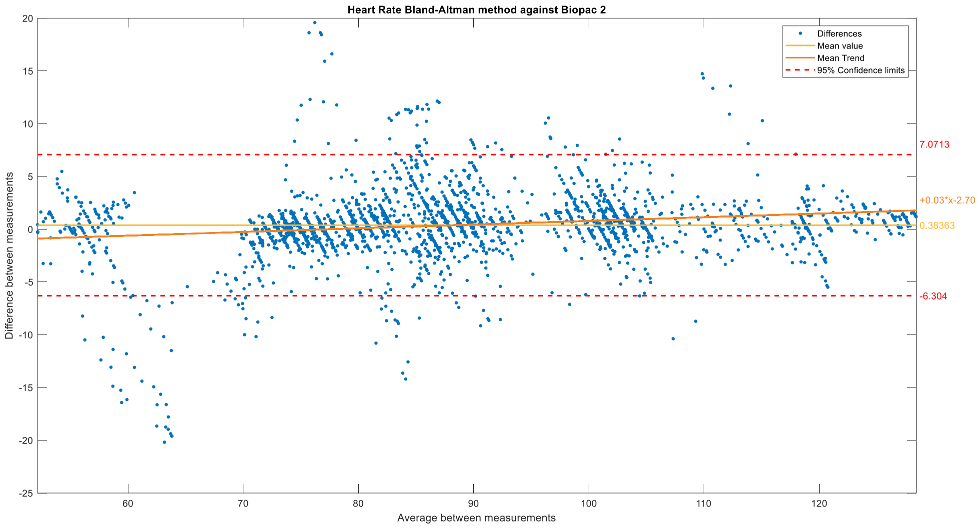

Figure 22 HR LifeVit measurements against Biopac 2 (Stand Up Position)

In heart rate measurements against Wellue oximeter it can be observed some differences. First of all, the overall mean value is reduced from ~0.80 to ~0.63 is not a big difference that can be attributed again to statistical variations in the sample. Another change is the value of mean trend slope because pass from negative to positive keeping the value of it. As a negative point, similar to the previous results in this position the confidence interval limits were increased, again due to some great outliers for low HR around 60 bpm.

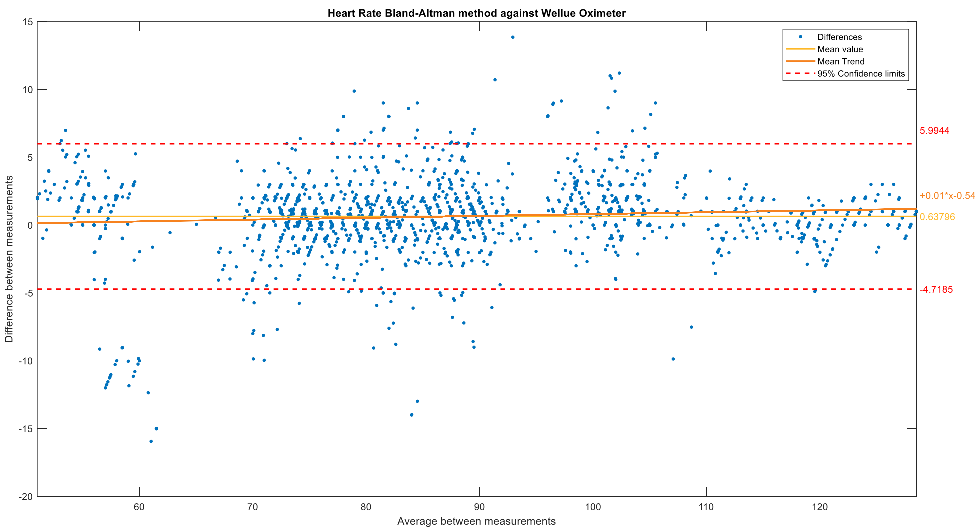

Figure 23 HR LifeVit measurements against Wellue oximeter (Stand Up Position)

3.1.5. Sport position

Finally, the Sports position is analysed. At the beginning was the position with more interest because is the most realistic if we think in the main application for DyCare but as the weeks went by and the first results were analysed it can be seen that it will be impossible to measure during rehabilitation exercises due to the incorrect working of LifeVit Vital in measuring HR and SpO2.

As in the previous cases, the heart rate measurements are better than the SpO2 measurements (see section 3.2.5), but in this case, compared to the actual values, both are too bad. It seems that in all evaluation methods in this step of the experiment the statistical values of all of them have been affected compared to the previous one.

Starting with Biopac results, at first glance can be seen that all the statistical values are out of position. First of all, the overall mean value increase from ~0.44 to ~8.88, in other words a 2018%, also the mean trend line can be observed, and it realizes that increase a 700% the slope value and the intercept increases a ~42000% its value. Also, the distance between the confidence interval limits increases a 1138%.

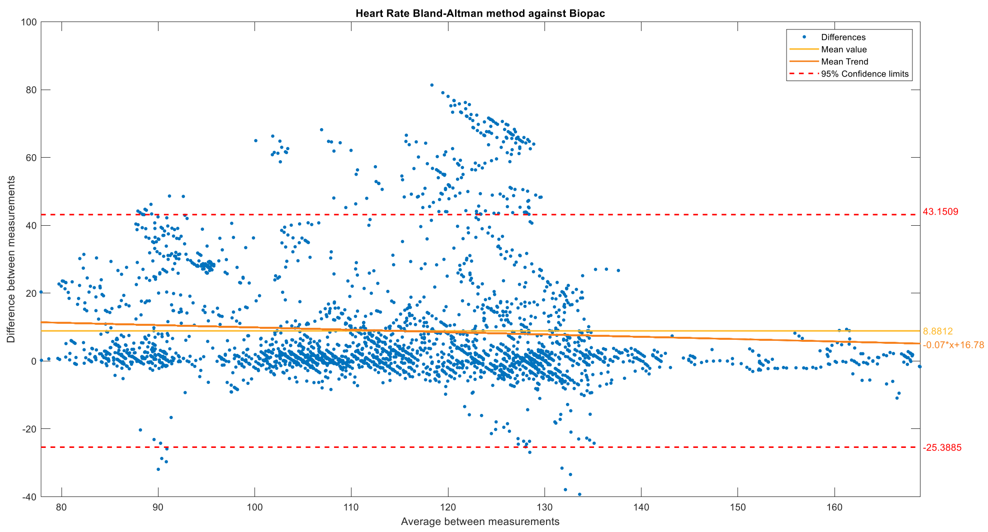

Figure 24 HR LifeVit measurements against Biopac (Sports Position)

Theses previous increases on the statistical parameters are critical for the future patients. It takes confidence out on the device results and can be risky for health monitoring. Looking to the Biopac 2 results (figure 25) were improved but nothing relevant as to increase the confidence.

It seems that all these bad results can have their origin in the technology used by the wearable in contrast with the Biopac that it uses electrodes that adhere to the skin and can measure properly the electrical behaviour of the heart while the wearable uses the previous commented photoplethysmography.

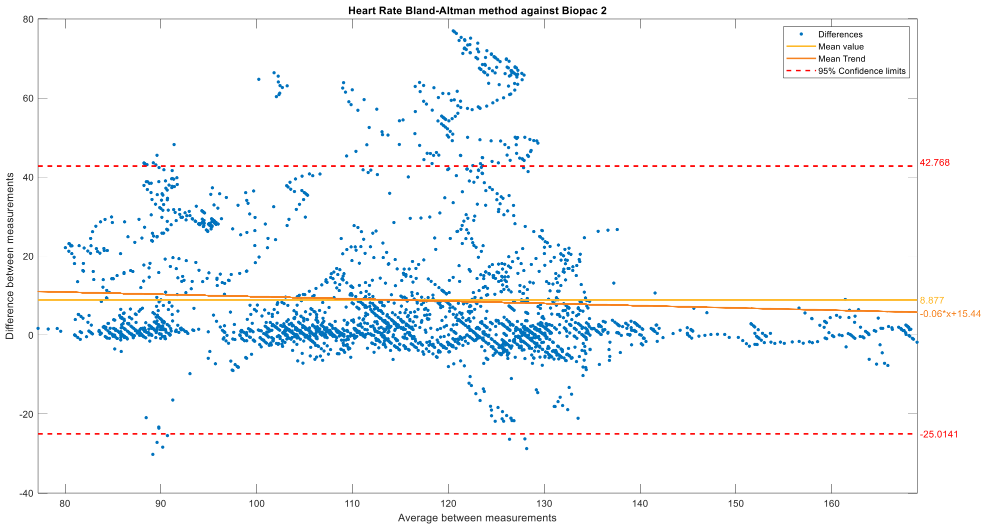

Figure 25 HR LifeVit measurements against Biopac 2 (Sports Position)

Analysing the measurements against Wellue oximeter again can be observed a superposition between the methods used by both. If the previous Biopac results were not good, in this case they are worst, i.e., the difference between confidence interval limits was increased compared with Biopac and the reference one.

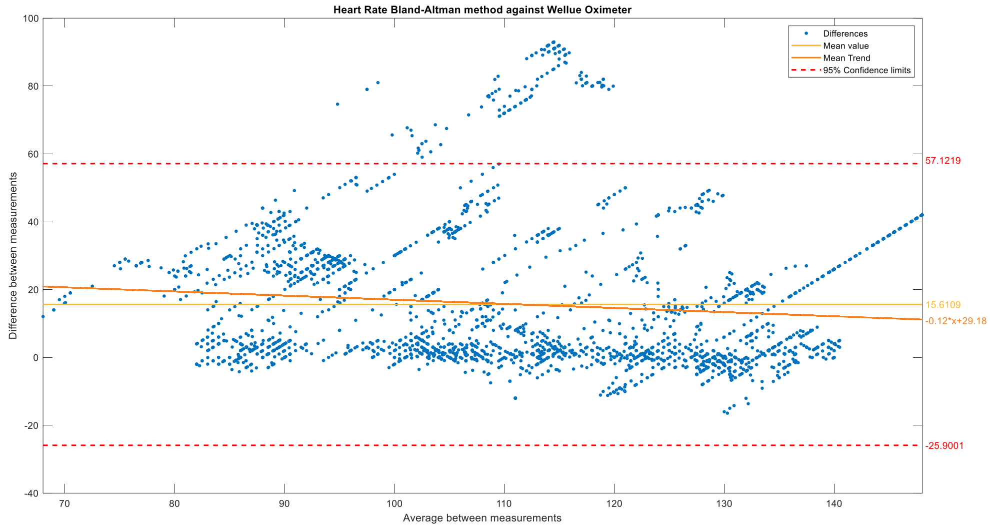

Figure 26 HR LifeVit measurements against Wellue oximeter (Sports Position)

3.2. SpO2 Results

Once, all the heart rate results have been properly analysed. The same order as in the heart rate will be used to help the correct understanding of the results. As it was commented in the previous section the device has problem with measurements in motion particularly when is doing SpO2 measurements, the device all the time returns “0” independently the value measured by the oximeter.

3.2.1. All Data

The results begin to be worst when LifeVit SpO2 measurements were compared with the reference oximeter. As is shown, it appears outliers again because of the sports situations when LifeVit wearable returns 0 independently of the measured value by the oximeter.

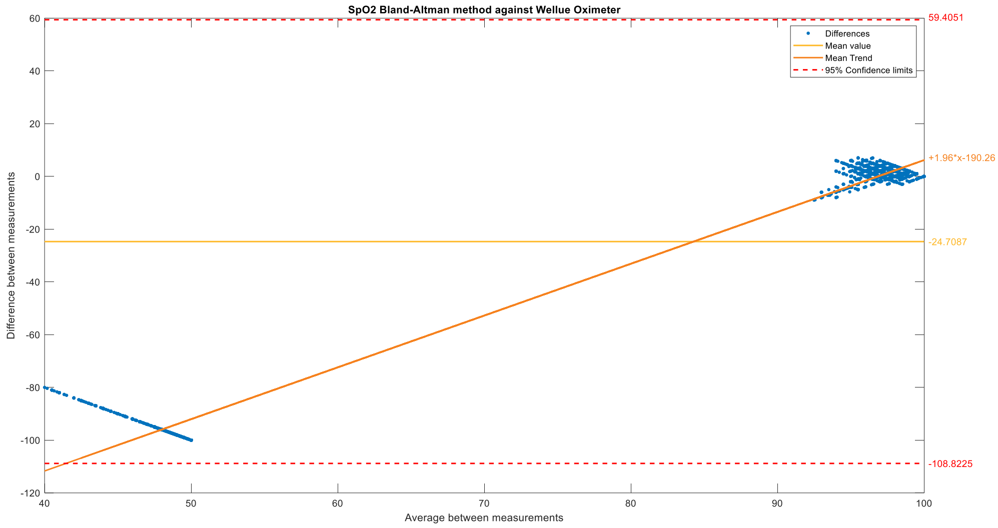

Figure 27 SpO2 LifeVit measurements against Wellue Oximeter

When comparing the measurements of the SpO2 with both devices without the sport situation are not as good as the heart rate measurements. This does not mean that the device is not measuring but due that the Spo2 range of normality is so reduced (between 95% to 100%) it can be critical if in some cases. It must be considered these results for the final application and conclusions.

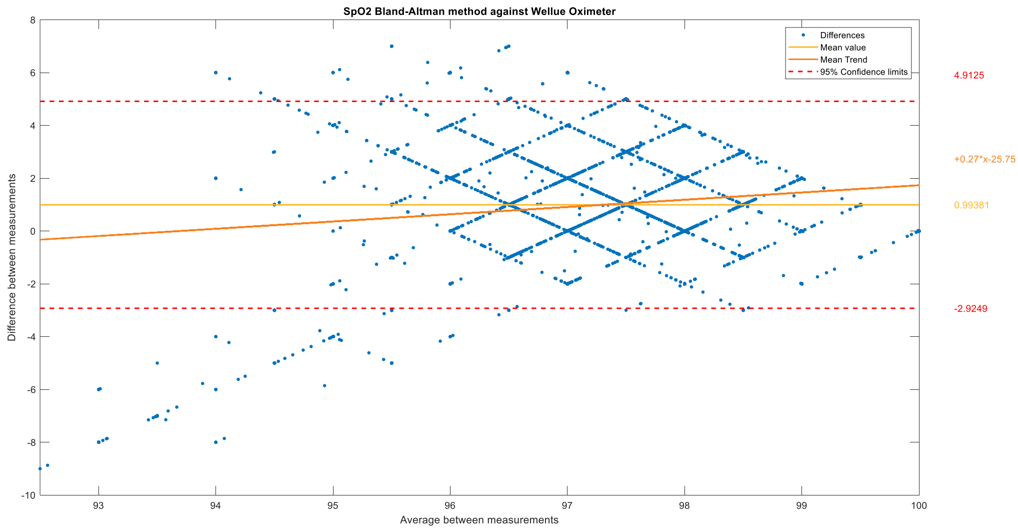

Figure 28 SpO2 LifeVit measurements against Wellue Oximeter without sport

3.2.2. Lie Down position

In figure 29 can be observed some similarities with the results in figure 28 and it is not good news because not only the overall mean was increased also the confidence interval limits were increased. In this case, is not normal due to during lie down position should not appear any movement, artefact or offset that can show us committed outcomes. The main reason of this results must be the outliers between ninety-three and ninety-five SpO2 measurements.

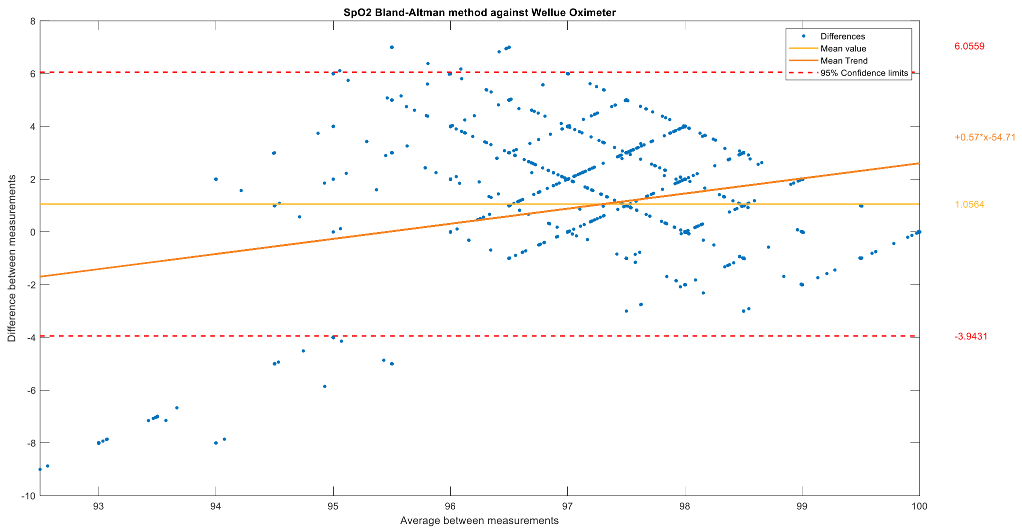

Figure 29 SpO2 LifeVit measurements against Wellue oximeter (Lie Down Position)

3.2.3. Sit Down position

The SpO2 measurements for sit down position, can be observed good news. All the statistical parameters were reduced. In the overall mean and confidence interval limit can be observed great differences. Both were reduced in a 50% against the reference position (lie down, see figure 29).

The linear regression results, orange line, was reduced compared with the lie down (figure 29) but suffers the same pattern. Again, we have a big gain error in SpO2 measurements but smaller then in lie down results.

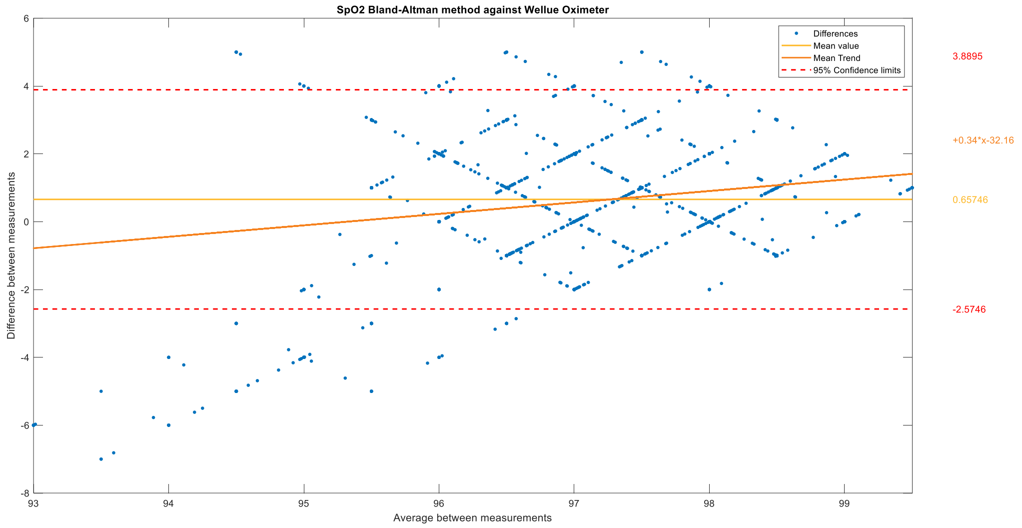

Figure 30 SpO2 LifeVit measurements against Wellue oximeter (Sit Down Position)

3.2.4. Stand Up position

In this position the SpO2 results are worse than the previous ones and the reference. Although the confidence interval limits are similar to previous ones, great differences can be observed in the overall mean value. The mean trend slope changed from positive to negative and represents a large gain error that reduces confidence, especially in SpO2 measurements that are more critical.

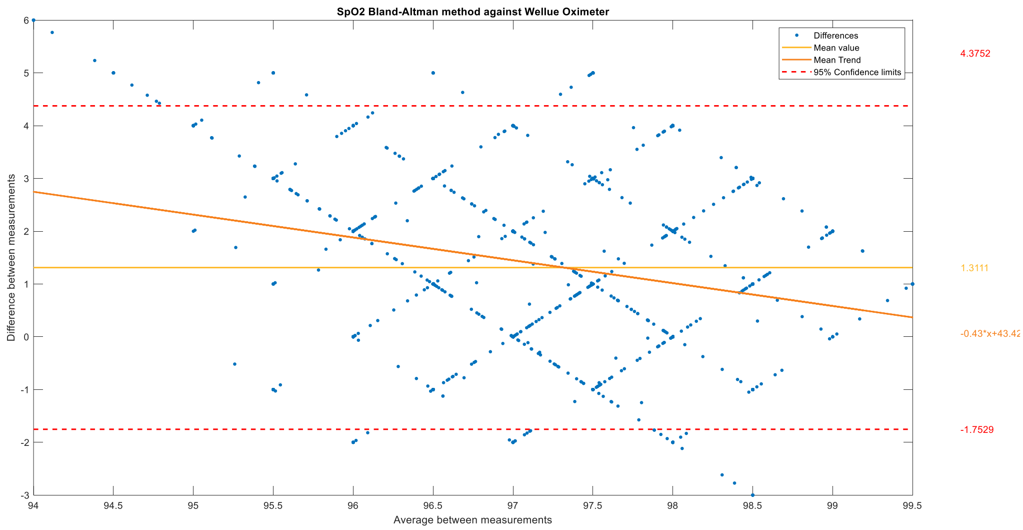

Figure 31 SpO2 LifeVit measurements against Wellue oximeter (Stand Up Position)

3.2.5. Sport position

Finally, the SpO2 measurements during a sports period were the worst and more critical compared with all previous ones. During the experiments the LifeVit returns always a 0 for the oximeter measurements (Figure 32).

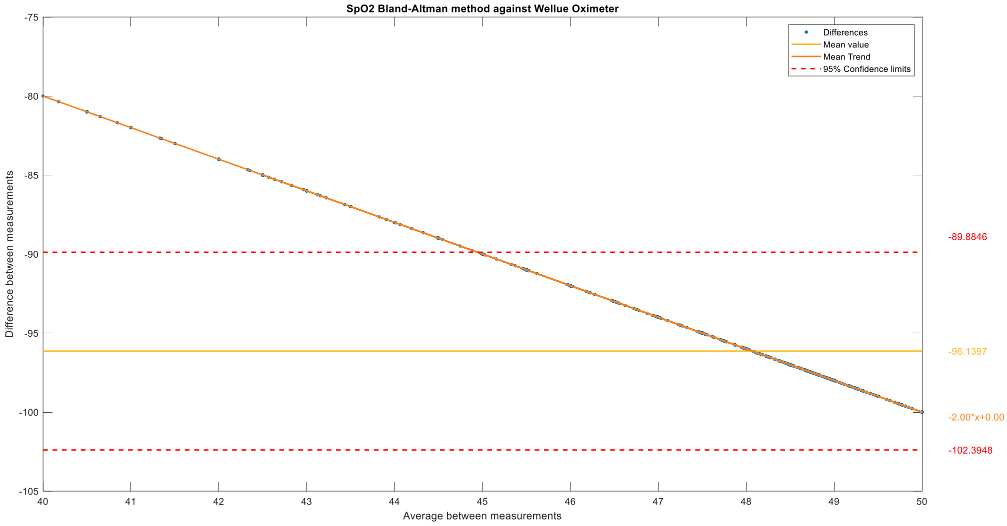

Figure 32 SpO2 LifeVit measurements against Wellue oximeter (Sports Position)

Finally, the following table shows the limits of agreement and mean results for the HR and SpO2 measurements in the four positions per device:

Position	Device	HR (bpm) mean	HR (bpm) LoA	SpO2 (%) mean	SpO2 (%) LoA
Lie Down	Biopac	0.44	-2.57 to 3.45	-	-
	Biopac 2	0.44	-2.24 to 3.13	-	-
	Wellue	0.87	-1.99 to 3.74	1.05	-3.94 to 6.05
Sit Down	Biopac	0.38	-4.54 to 5.32	-	-
	Biopac 2	0.38	-4.02 to 4.79	-	-
	Wellue	0.80	-3.20 to 4.82	0.65	-2.57 to 3.88
Stand up	Biopac	0.38	-6.96 to 7.73	-	-
	Biopac 2	0.38	-6.30 to 7.07	-	-
	Wellue	0.63	-4.71 to 5.99	1.31	-1.75 to 4.37
Sport	Biopac	8.88	-25.38 to 43.15	-	-
	Biopac 2	8.87	-25.01 to 42.76	-	-
	Wellue	15.61	-25.90 to 57.12	-96.13	-102.39 to -89.88

Table 3 Summary table results

4. DISCUSSION, CONCLUSION and SUGGESTIONS

Consequently, once all the previous results have been presented and deeply analysed can be extracted some conclusions and answered some questions related with the purpose of the thesis. Also, it must be said that all these conclusions, suggestions and answers are personal opinions, and the previous results may be interpreted in different ways by other professional researchers.

Firstly, can be said that the wearable does not have safety problems but must be considered that previously to be used it must be properly adjusted. Some considerations to this well-adjustment could be:

- The device must be wearing 1-2 fingers above the pisiform bone.
- If the volunteer has dark skin the devices could have some problems to measure properly.
- To measure the heart rate, if the device is adjusted a few minutes before the measurements, the results are displayed immediately instead of taking a few seconds to obtain them The movement can produce non-realistic values.
- For SpO2 readings it takes some seconds to obtain normal measurements, meanwhile it sends only zeros, and it has directly correlated with the movement, any small movement produces not expected zeros.

Secondly, in accuracy terms it seems that it depends directly on the biological signal to measure. In heart rate, the device is offer reasonable results, and it did not have any problem to measure in most of the situations. As can be observed on the previous section, the best results were for lie down position, it is true that in stand up and sit down positions were obtained a less mean value than in previous one but that only means that in the whole experiment the difference between samples was minor. However, it is important to mention that the difference between confidence interval limits was minor in the first position than the other positions, and we did not observed outliers during the experiment.

Another aspect that was observed was that the volunteers #0 and #11 obtained the best results. To see how good these results were, can be seen an example in section 2.1.4.

As well as that, the code developed in Matlab to interact with the device worked correctly. In principle, it must not be expected any risks or issues when the device will be integrated in ReHub. Once the device will work properly with the platform can be done an extra assess in real time comparing the values obtained against an oximeter to be 100% sure that the results obtained can have some confidence, but it must not see any difference. Another solution can be correcting the measurements with the linear regression results obtained during the Bland-Altman method.

To summarize, the wearable LifeVit Vital can be incorporate to the ReHub telerehabilitation platform with the confidence in heart rate measurements is effective but considering that the SpO2 values can be far away from the real one. It is true that SpO2 results were not good but applying some considerations on the platform and informing the patient about the risks is better than nothing.

Considering the patients side, they should be informed with all the information found in the results and let them know that without this device they have more risks that not using it and remarking that this is not a medical device but a solution to help the professional supervising their health during their exercises in case that they suffer any problem. Once the device will be on the market can be extracted more data to assess if with new firmware updates or algorithms the results can be improved comparing with a reference oximeter as it was used during the experiments.

5. DEDICATION

Absolutely this will be the most informal thesis part, but it is necessary. As all my previous professional and academic achievements, I want to dedicate this thesis to my mum because, alone since my father is not with us, in the worst moments of my life she always reminds me that always it is necessary to keep calm, keep your head up and face this new challenge that life has set to you.

Thanks mum.

BIBLIOGRAPHY

1. *EUR-Lex - 32017R0745 - EN - EUR-Lex*. (n.d.-b). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745
2. Bayoumy, K., Gaber, M. A., Elshafeey, A., Mhaimeed, O., Hill, E., Marvel, F. A., Martin, S. S., Muse, E. D., Turakhia, M. P., Tarakji, K. G., & Elshazly, M. B. (2021). Smart wearable devices in cardiovascular care: where we are and how to move forward. *Nature Reviews Cardiology*, *18*(8), 581–599. https://doi.org/10.1038/s41569-021-00522-7
3. *Accurate Detection of Heart Rate and Blood Oxygen Saturation in Reflective Photoplethysmography*. (2020, December 9). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9408845
4. Dias, D., & Cunha, J. P. (2018). Wearable Health Devices—Vital Sign Monitoring, Systems and Technologies. *Sensors*, *18*(8), 2414. https://doi.org/10.3390/s18082414
5. Alfonso, C., Garcia-Gonzalez, M. A., Parrado, E., Gil-Rojas, J., Ramos-Castro, J., & Capdevila, L. (2022). Agreement between two photoplethysmography-based wearable devices for monitoring heart rate during different physical activity situations: a new analysis methodology. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-18356-9
6. Altman, D. G., & Bland, J. M. (1983). Measurement in Medicine: The Analysis of Method Comparison Studies. *Journal of the Royal Statistical Society. Series D (The Statistician), 32(3), 307–317.* https://doi.org/10.2307/2987937
7. Gerke, O. (2020). Nonparametric Limits of Agreement in Method Comparison Studies: A Simulation Study on Extreme Quantile Estimation. *International Journal of Environmental Research and Public Health*, *17*(22), 8330. https://doi.org/10.3390/ijerph17228330
8. *A Real-Time QRS Detection Algorithm*. (1985, March 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/4122029
9. *1-D digital filter - MATLAB filter*. (n.d.). https://www.mathworks.com/help/matlab/ref/filter.html?s_tid=doc_ta

10. *Zero-phase digital filtering - MATLAB filtfilt*. (n.d.). https://www.mathworks.com/help/signal/ref/filtfilt.html
11. Spaccarotella, C., Polimeni, A., Mancuso, C., Pelaia, G., Esposito, G., & Indolfi, C. (2022). Assessment of Non-Invasive Measurements of Oxygen Saturation and Heart Rate with an Apple Smartwatch: Comparison with a Standard Pulse Oximeter. Journal of Clinical Medicine, 11(6), 1467. https://doi.org/10.3390/jcm11061467