A statistical analysis of the viability of heart rate detection for sleep apnea using film BCG

Aly Khaled **1190156** , Mohamed Nasser **1190438 ,** Mariam Aly **1190519**, Maryam Moataz **1190522**

Cairo University, Faculty of Engineering
SBEN429: Data Analytics

Table of Contents

Table of Figures

1. Abstract

Contactless, remote monitoring of vital signs (VS) is an emerging technology with an escalating demand and a vast number of possible uses such that it can help increase the number of seniors living alone in their homes and aid caregivers. Heart Rate is one of the most fundamental vitals that need monitoring at all times, even during sleep. But wearing wearable devices isn't always the best option and can have drawbacks especially with elderly people. This is where contactless remote monitoring comes in handy. Like any new idea, testing and assessment is very important, which was the goal of this project.

2.Introduction

Achievement of accurate and reliable diagnosis and treatment of sleep apnea (SA) is challenging due to the complexity of the illness and its associated symptoms. While polysomnography (PSG) remains the gold standard for diagnosing SA, its cost and time-intensive approach make it difficult to use in a clinical setting. As such, new technologies, such as the use of electrocardiography (ECG) and ballistocardiography (BCG), have been developed to address some of the shortcomings of PSG. This article aims to discuss the differences between ECG and BCG for heart rate detection in the context of SA.

ECG vs BCG

The electrocardiogram (ECG) is a standard tool used in clinical settings to detect the electrical activity of the heart. It uses electrodes placed on the body to measure the electrical signals that cause the heart's muscles to contract and expand. A ballistocardiogram (BCG) is a recording of the mechanical forces generated by the heart during each heartbeat, such as the amount of force generated against the chest wall. The ECG is more commonly used in medical settings to diagnose problems, while the BCG is more commonly used in research settings.
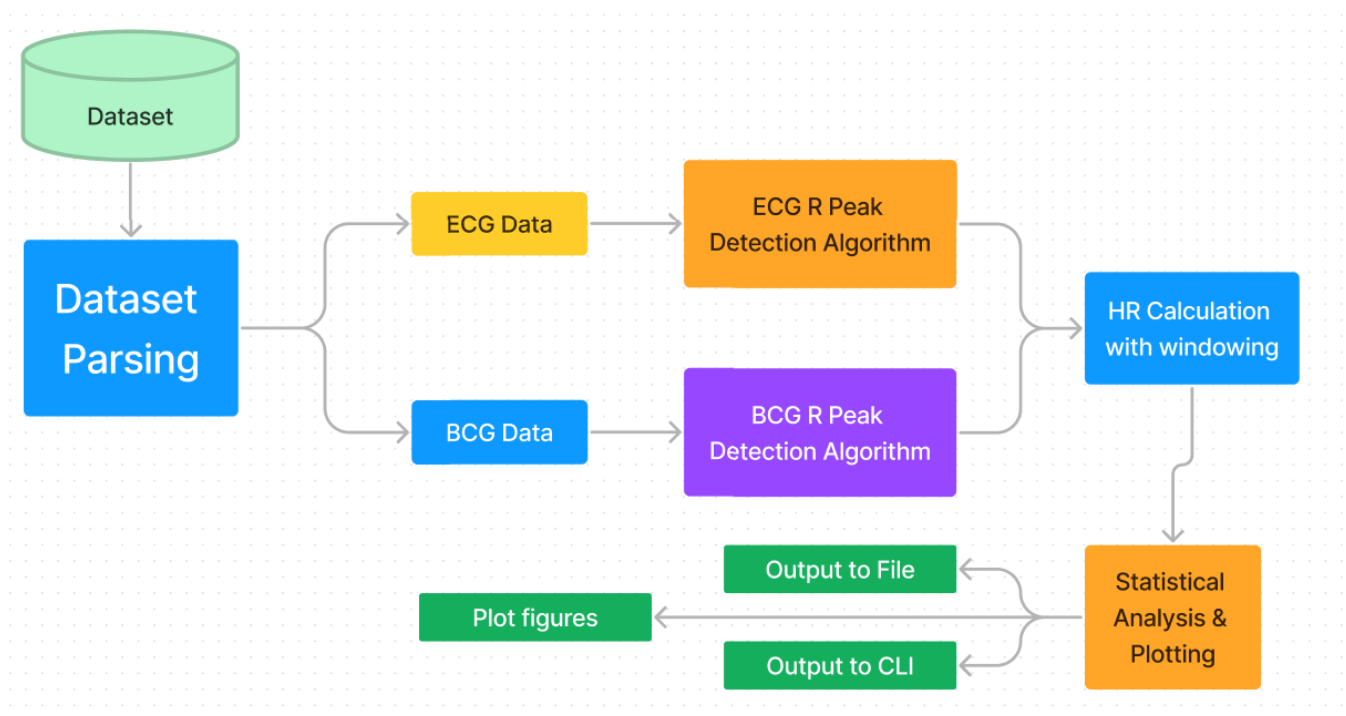
3. Program Block Diagram

Figure 1 Block diagram of program

4.Statistical Analysis Plots

We had to compute three different plots; Bland-Altman plot, boxplot, Pearson Correlation plot between estimated HR and reference HR. Let's briefly review these plots.

- ***The Bland-Altman*** plot, commonly known as the difference plot, is a useful display of the relationship between two paired variables using the same scale. Figure (5,8) shows the Bland-Altman plot that we computed for on patient and for all patients.

- ***The boxplot*** is a standardised way of displaying the distribution of data based on a five number summary ("minimum", first quartile [Q1], median, third quartile [Q3] and "maximum"). . Figures (4,7) show the ECG and BCG boxplots that we computed for the same patient, and the whole dataset.

- ***Pearson Correlation*** A positive correlation signifies that if variable A i.e. [ECG signal] goes up, then BCG signal will also go up, whereas if the value of the correlation is negative, then if ECG signal vitals increases, BCG signal vitals decreases.

  The Pearson Correlation coefficient can be computed in Python using **corrcoef()** method from numpy: The input for this function is typically a matrix, say of size mxn, where:

  1-Each column represents the values of a random variable

  2-Each row represents a single sample of **n** random variables

  3-**n** represent the total number of different random variables

  4-**m** represents the total number of samples for each variable

Used Built-in Functions

  .scatter(); is used to plot the output matrix

  .sns.heatmap(); is used so that they replace numbers with colours of varying shades, as indicated by a scale on the right.
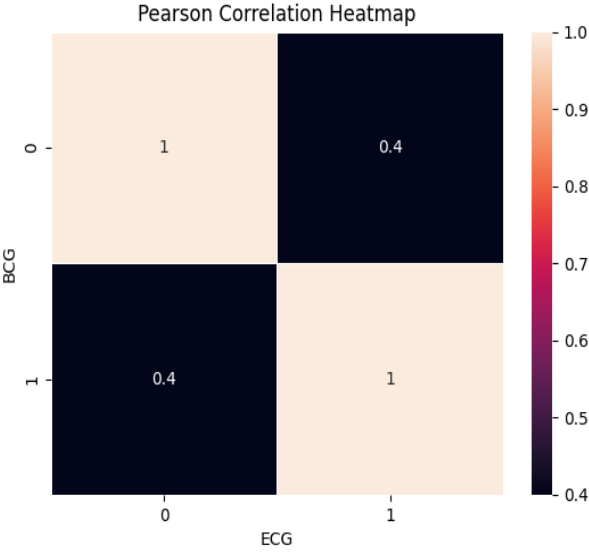
● *Plots for 1 Patient*

Figure 2 Pearson correlation heatmap for one patient
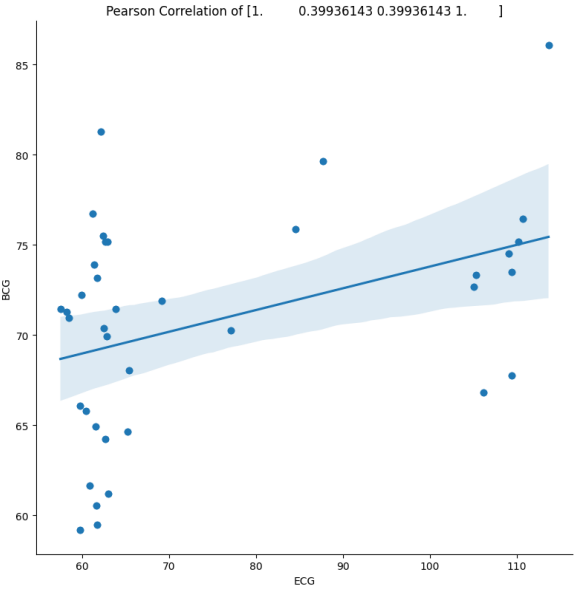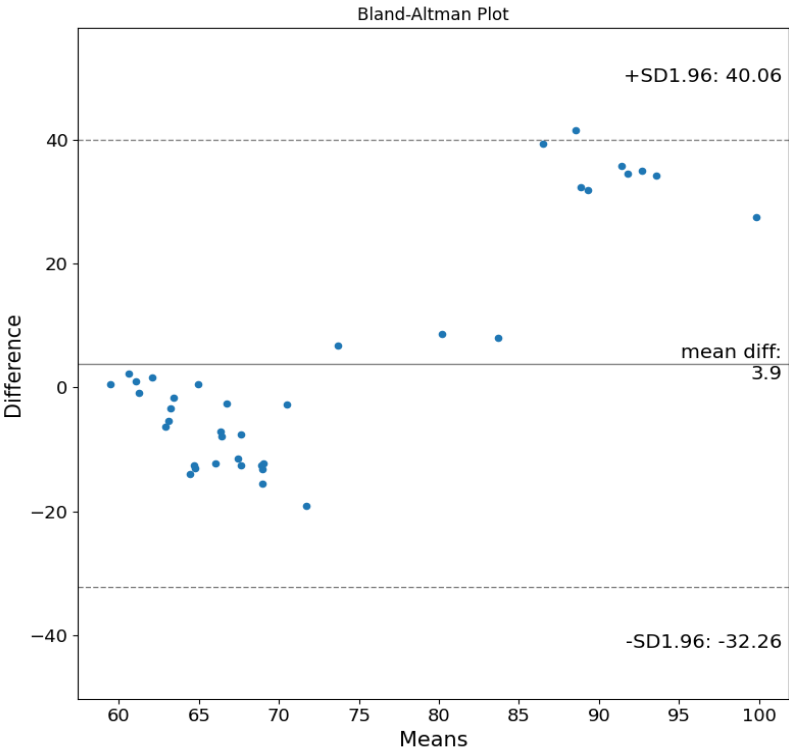

Figure 3 Pearson correlation plot for one patient

Figure 5 Bland–Altman plot for one patient
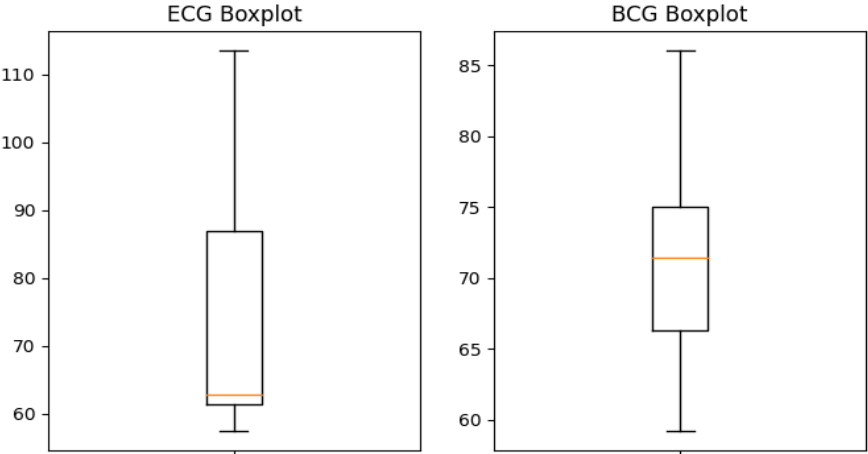
Figure 4 Box plots for BCG and ECG signals, one patient

● *Plots for All patients*

Bland-Altman Plot

+SD1.96: 15.22

mean diff: −1.85

-SD1.96: -18.91
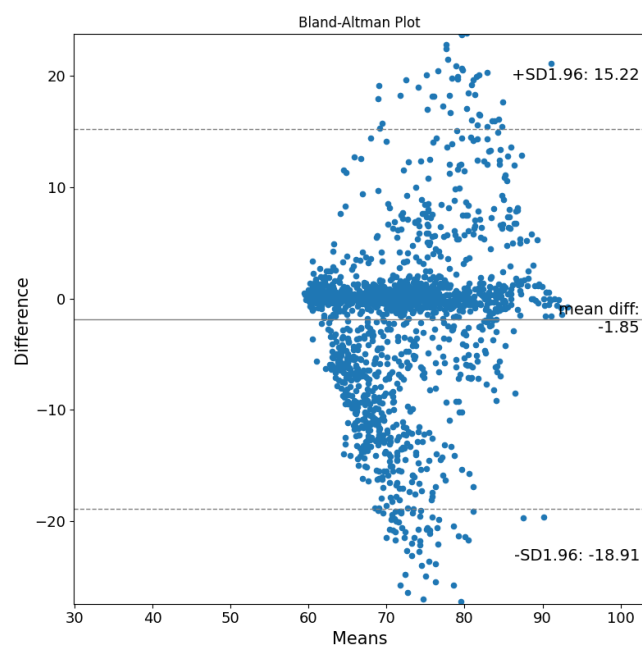
Figure 8 Bland–Altman plot for 40 patients

Pearson Correlation of [1.   0.49458824 0.49458824 1.   ]
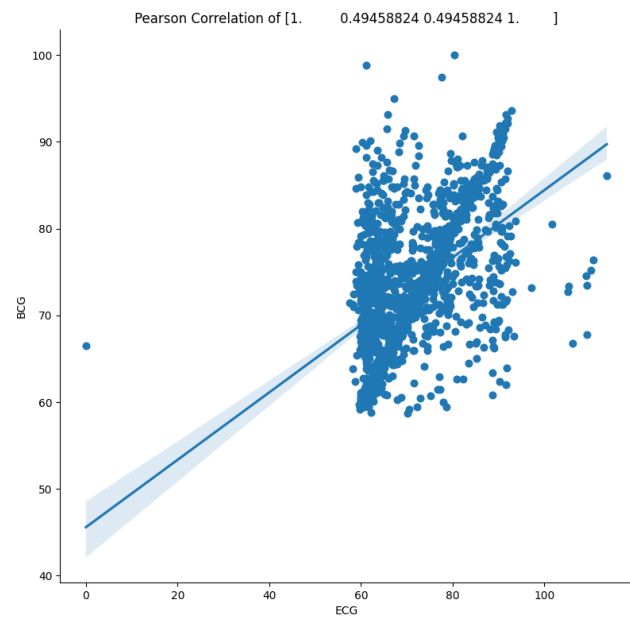
Figure 6 Pearson correlation plot for 40 patients

ECG Boxplot   BCG Boxplot

Figure 7 Box plots for BCG and ECG signals , 40 patients

5. Vitals Statistics

We also had to calculate the **mean absolute error, root mean square error,** and **mean absolute percentage erro**r for each patient, but what is the significance of calculating these errors?
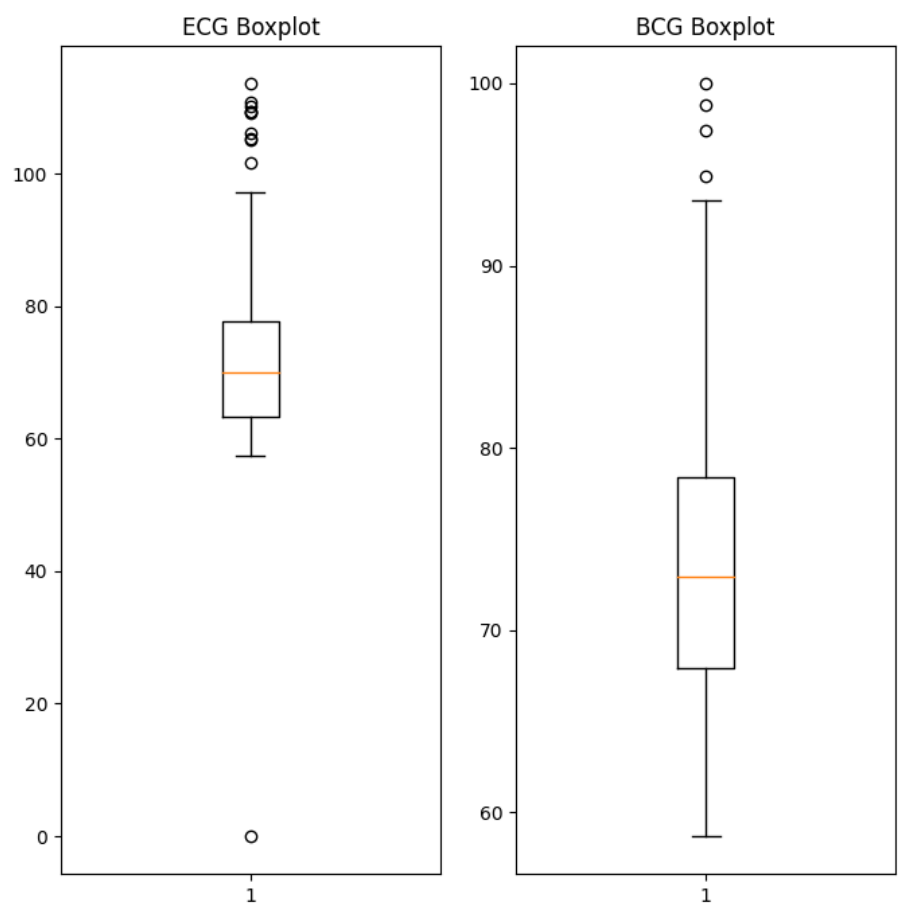
- **The Mean Absolute Error** is a measure of errors between paired observations expressing the same phenomenon. It measures accuracy for continuous variables which are the ECG and BCG in our project by showing the magnitude of difference between the prediction of an observation and the true value of that observation.

- **The Root Mean Square** error measures the differences between values (sample or population values) predicted by a model or an estimator and the values observed. It provides an estimation of how well the model can predict the target value (accuracy) which is the HR in our project.

- **The Mean Absolute Percentage Error** is the mean or average of the absolute percentage errors of forecasts. Error is defined as actual (ECG) or observed value minus the forecasted value (BCG), divided by the actual values (ECG) for each period.

The following text snippet shows each of the Mean Absolute Error, Root Mean Square and Mean Absolute Percentage Error for a patient .

```
---------------------
patient id : 1008
Heart Rate Information
---------------------
ECG Method
Minimum pulse : 57.0
Maximum pulse : 114.0
Average pulse : 75.0
---------------------
BCG Method
Minimum pulse : 59.0
Maximum pulse : 86.0
Average pulse : 71.0
---------------------

Plotting results...
-----------ERROR CALCULATIONS FOR PATIENT WITH ID : 1008-----------
Mean Absolute Error = 14.113646458260007
Root Mean Square Error = 18.856824608591534
Mean Absolute Percentage Error = 16.92859039698412%
```

6.Problems We Faced

I.   Converting the dataset

To be able to read the study's dataset we first had to convert it from .mat format to comma separated values (.csv). To do this we had to write matlab code that loops through the datatable and parses it accordingly, separating the 40 patients' data.

This is a snippet of the code we used

```
1    FileData = load('Preprocessed_Database.mat');
2    data = FileData.Preprocessed_Database;
3    cleanData = removevars(data,{'FilteredData'});
4    rawDataCol = [""];
5
6    for i = 1:40
7        rawDataCol(i) = string(cleanData{i,1})+".csv";
8
9    end
10   cleanData.RawData = rawDataCol.';
11   writetable(cleanData,'myData.csv','Delimiter',',','QuoteStrings',true)
12
13   for i = 1:40
14       rawData = data{i,2}{1,1};
15       writetable(rawData,'filtererdData/'+rawDataCol(i),'Delimiter',',','QuoteStrings',true)
16       disp("Done")
17   end
```

II.   Analysing the ECG signal

The signal was first downsampled from (1KHz) to (50Hz) then processed with discrete wavelet transform using the biorthogonal wavelet 3.9 then *heartpy* was used to count the peaks and calculate the HR from the signal using **hp.process_segmentwise()**
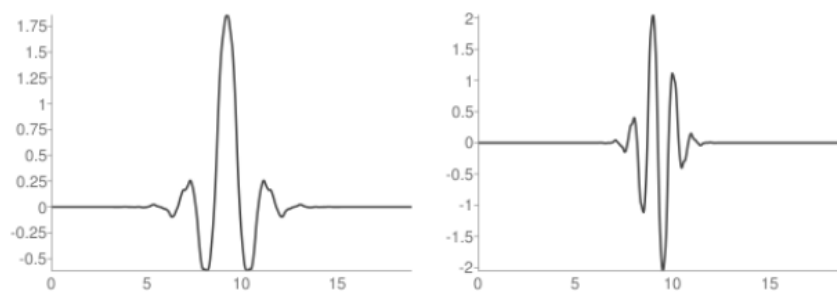

Figure 9 Bio orthogonal wavelet 3.9

III. Adapting the original code

1. Removed Breathing Signal Analysis, Bandpass filtering step
2. Modified Vitals func to calculate heart rate using sampling frequency instead of time axis

```python
6   def compute_rate(beats,mpd,fs):
7
8       indices = detect_peaks(beats, mpd=mpd)
9
10      if len(indices) > 1:
11          diff_sample = indices[-1] - indices[0] + 1
12          t_N = diff_sample / fs
13          heartRate = (len(indices) - 1) / t_N * 60
14          return heartRate, indices
15      else:
16          return 0.0, 0.0
```

IV. Output Error Handling

Some patients would inadvertently cause a bcg, ecg array length mismatch due to slightly different signal lengths, and zero division or wavelet fitting errors that would cause NaN signal outputs. To handle these errors conditions were added as follows:

```python
77          # if nan in hr_ecg or nan in hr_bcg remove it
78          if np.isnan(hr_ecg).any():
79              hr_ecg = np.nan_to_num(hr_ecg)
80          if np.isnan(hr_bcg).any():
81              hr_bcg = np.nan_to_num(hr_bcg)
82
83          # if one is larger than the other, remove the last element
84          if len(hr_ecg) > len(hr_bcg):
85              ecg_hr = hr_ecg[:-1]
86          elif len(hr_ecg) < len(hr_bcg):
87              hr_bcg = hr_bcg[:-1]
```

7. Conclusion

According to the statistical analysis performed on multiple patients it is apparent from the bland-altman plot that the mean difference is within acceptable margins, and so is the standard deviation. This is indicative of the accuracy of the BCG and renders it as a promising device for future household use to help many of those who suffer from sleep apnea.

References

Sadek, Ibrahim, and Bessam Abdulrazak. "A comparison of three heart rate detection algorithms over ballistocardiogram signals." Biomedical Signal Processing and Control (2021).

7. Correlation and Scatterplots — Basic Analytics in Python (sfu.ca)

Sadek, Ibrahim, et al. "A new approach for detecting sleep apnea using a contactless bed sensor: Comparison study." Journal of medical Internet research
(2020).

NumPy, SciPy, and Pandas: Correlation With Python – Real Python