Journal of Engineering Sciences
Vol 10, Issue 12, DEC/2019
ISSN NO:0377-9254

MACHINE LEARNING BASED SYSTOLIC AND DIASTOLIC PRESSURE MONITORING

#1T.DEEPIKA, M.Tech in DIGITAL SYSTEMS AND COMPUTER ELECTRONICS,
#2 Mr.P.RAJESH, Assistant Professor [Adhoc],
#3 Dr.D.VISHNU VARDHAN, Assistant Professor
#1,2,3DEPT OF ECE JNTUA, ANANTHAPURAM

ABSTRAT: Continuous BP estimation is utilized to recognize the potential danger of cardiovascular sickness and give precious data to clinical analysis. This work presents an efficient method based on the Pulse transit time. Here we use Electrocardiograph (ECG) and Photoplethysmograph (PPG) for nonstop and cuff less estimation of systolic pulse (SBP) and Diastolic blood pressure (DBP). For the extraction of features from ECG signal we use spectral masking approach along with wavelet and for PPG signal the principal of first derivative is used to extract the notch region and other parameters such as peak and valley. Finally the regression algorithms are employed for the estimation of BP.

Keywords: Photoplethysmograph (PPG), Electrocardiograph (ECG), Pulse Transit Time (PTT)

I. INTRODUCTION

Based on the information given by the World Health Organization (WHO), hypertension pervasiveness is 24 percent in men and 20.5 percent in women [1]. But vast majority of patients don't know about their hypertension, which causes damage to the internal organs of the body. For exact analysis and treatment of hypertension, BP estimation is vital. BP is a periodic signal with the Heart Rate (HR) recurrence that is ordinarily limited in a constrained range. The BP swings among greatest and least qualities, which are called Systolic blood

pressure (SBP) and Diastolic Blood Pressure (DBP), individually. In the past few years there has been a great attention regarding cuff less BP estimation strategies toward the aim of constant BP monitoring. Using the Pulse Wave Velocity (PWV) is the most widely recognized technique [2]. The velocity of the pressure wave which is propagating through the vessel is known as PWV. This technique mainly depends on wave propagation theory of fluids in elastic pipes. Pulse Transit Time (PTT), the time interval taken by a heartbeat to travel to the body peripherals is used for the estimation of PWV. Ahmad et al. [3] and Xuan et al. [4] demonstrated that there is a huge correlation between BP and PTT but this correlation is dependent on many parameters which changes from person to person. Numerous other works attempted to fit regression models for BP estimation utilizing PTT, yet did not fulfill the standard criteria. The correlation between's people's BP and their PTT is calculated in various circumstances. They examined the connection between the ordinary subject's BP and PTT in six-month time span. Gesche et al. [5] recommended a calibration technique for the estimation of dependency. But such calibration is dependable only for a short interval of time. In spite of the fact calibration based techniques cannot be dependably utilized as substitutions of the conventional BP estimation tools, they are relevant for BP checking in short time intervals such as exercise tests.

Regardless of their benefits, the PWV based strategies are confronted with a few troubles, for example, reliance to person physiological parameters, which requires detailed calibration techniques. This concern prevents the PWV based estimation strategies to be approved by the health care standards. Thus, the PWV based techniques practically can't replace the regular BP estimation strategies.

This work exhibits a novel methodology that utilizes different signal processing and machine learning algorithms to accomplish an exact and persistent estimation of BP in healthcare monitoring systems. In

synopsis, after removing the noise from ECG and PPG signals the features are extracted and finally these features are given as the inputs to the regression models which estimate blood pressure.

II. LITERATURE REVIEW

The vascular system can be considered as the elastic tubes in which blood flows. The brief information about the blood flow and the properties of the arterial walls are discussed below.

The arterial wall is constructed by four layers called endothelium, elastin, collagen, smooth muscle. Endothelium is a layer which acts as the wall for the flow of the blood in the vessel. Elastin has considerable elastic properties which can produce tension on the arterial walls. Collagen is a layer with more stiffness compared to the elastin. It applies stress as soon as the wall of the arteries is stretched. Because of its properties, elastin is responsible for the flexibility in blood vessel at low BP values; while collagen fundamentally decides the blood vessel elasticity at high BP values.

Here for the calculation of the Blood Pressure we used the ECG signal as the proximal end and the PPG signal as the distal end. The time taken by the pressure wave to travel from proximal end to the distal end in an arterial wall is considered as the PTT. This pulse transit time is used in the estimation of the pulse wave velocity. In previous methods Pulse Arrival Time which is defined as the time gap between the heart's electrical activation and the pressure pulse arrival at the distal point, can be used as the measure of PWV. It is shown that using PAT can reduce the diastolic blood pressure accuracy. So instead of pulse arrival time the pulse transit time is used for the estimation of pulse wave velocity. In order to measure the PTT value various vital signals like seismocardiogram (SCG), ballistocardiogram(BCG), electrocardiogram (ECG), and photoplethysmograph (PPG) [6] can be used. But for the machine learning purpose which requires large number of data the ECG and PPG signals are used.

III. PROPOSED METHOD

This section involves four modules to describe the proposed methodology.

Module 1: Database

In this work, the ECG and PPG signals as well as ABP signal are taken from the database known as Physionet Multi parameter intelligent monitoring in intensive care (MIMIC II) Database [7]. Physionet database is website which contains patient's medical records in value and waveform format. The ABP signal is used to calculate the Systolic and Diastolic Blood pressure values.

Module 2: ECG & PPG signal processing

The Module-2 describes processing of ECG and PPG signal. The algorithm which is used for the signal processing is shown below.

Algorithm 1:

1	Notch filter implementation using H(z)
2	Baseline removal using db6 as mother wavelet
3	Subtracting approximation signal from the original signal.
4	Smoothing signal to nullify glitches if present.

2.1 Notch Filter Implementation:

Notch filter is implemented using H (z) as shown in (1). The filter implementation is carried out for 50Hz line interference, with the sampling frequency of 500 Hz.

$$H(z) = \frac{0.7548\, s^2 - 1.221\, s + 0.7548}{s^2 - 1.221\, s + 0.5095} \quad \text{.... (1)}$$

2.2 Base-line Wander removal & Smoothing filter:

Due to the Breathing effect there is a motion artifact which influences to cause Base line wander, Presence of glitches can be removed with the use of smoothing filter. Smoothening filter is created by using average filter.

2.3 Peak Detection:

This is the most critical step for any software performing Heart Rate Variability. There has to be an accurate and robust detection algorithm for R-R interval estimation. Peak detection[8] is carried out in two steps where in the beginning it is carried out as an automated process and secondly as a manual process. The equations (2) and (3) would derive the peaks of ECG signal and in manual process all the points which are not detected are processed again with a variation in threshold.

2.4 Peak to Peak Interval and Heart Rate Calculation:

This algorithm evaluates peak over the complete sequence of data and then sets the difference of data into a temporary variable as RR interval. The i-th RR interval can be obtained by the difference between the consecutive R waves as shown in (2).

$$RR_i = t_i - t_{i-1}. \qquad -(2)$$

$$RR_{avg} = \frac{1}{N}\sum_{i=1}^{N} RR_i \qquad -(3)$$

This RR interval provides a time series which can be shown as a function of time, as (t_{in},RR_i).

Average of R-R interval is shown in (3) where N represents total number of RR interval. It then calculates the heart rate.

Module 3: Feature extraction

The time domain features are extracted from the RR intervals such as variation in heart rate which is given as Heart rate variability analysis. The HRV analysis gives the basic information about the changes in the interval of time between the successive heart beats. The most commonly used method to analyze HRV is a time domain method called RMSSD. The RMSSD is defined as the root mean square of successive differences between each heart rate and is calculated by using the formula

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2} \qquad ...(4)$$

The most common method which is used to visualize the HRV is a graph called the Poincare plot. It is a graph in which each RR interval is plotted as a function of preceding RR interval where the data of every consecutive RR interval define a point in the plot. Pulse Transit Time: PTT gives the information about the time taken by the pressure waveform to travel through the length of arterial tree. The PTT is calculated by subtracting the ECG R peak time from the PPG peak time.

Pulse wave velocity: PWV is the velocity of the blood pressure pulse which travels through the entire circulatory system. Pulse wave velocity is calculated by using the pulse transit time and the equation is given below.

$$PWV = \{L / mean(PTT)\}/100 \qquad ....(5)$$

L= Distance between suprasternal notch and the femoral site. The PWV plot is shown in figure.

Module 4: Regression Models

In the formation of dataset for the supervised learning task the ABP signal is used to obtain the resultant SBP and DBP values. With the help of features which have been extracted from ECG and PPG signals for the estimation of SBP and DBP values are trained here. Here multiple linear regression models are used to perform the regression task.

4.1 Multiple Linear Regression (MLR):

Multiple Linear Regression (MLR), which is also referred as Multiple Regression [9] is a procedure that uses a few illustrative factors to estimate the result of the reaction variable. The objective of multiple linear regression (MLR) is to demonstrate the linear connection between the informative (independent) factors and reaction (dependent) variables.

Fundamentally, multiple regression is the extension of ordinary least-squares (OLS) regression that includes more than one illustrative variable.

Formula for multiple linear regression is given as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon$$

Where i = n observations

Yi = Dependent variable

Xi = Explanatory variables

βo = Y-intercept (constant term)

βp = slope coefficient for each explanatory variable

ϵ = Error term

4.1.1 Leverage: In Regression analysis how far the independent variable estimations of an observation are from different observations. The leverage plots are utilized to decide the impact of data points on the test for incorporating the impact in the model. The case order plot of leverage for SBP and DBP are shown in Fig 1

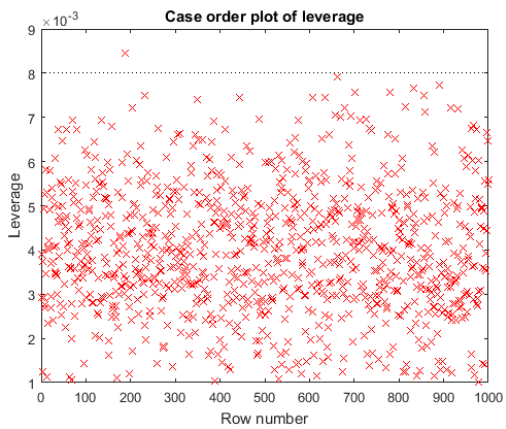


Fig (a)

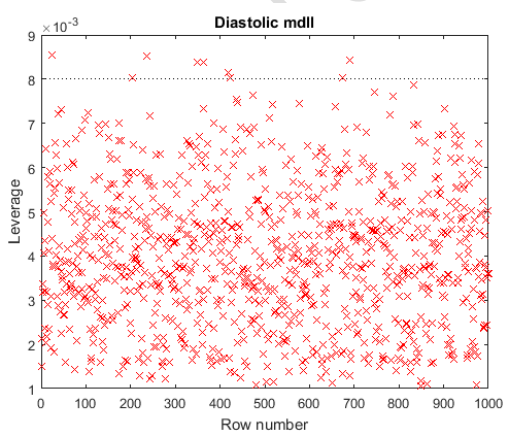


Fig (b)

Fig1. Leverage plot of a) SBP and b) DBP

4.1.2 Cook's Distance: The measure of changes in regression coefficients when an observation is deleted is defined as the Cook's distance. Information with large residuals (outliers) and additionally high leverage may manipulate the result and precision of a regression. Cook's Distance estimates the impact of erasing a given observation. Data points with an enormous Cook's distance are considered to justify nearer assessment in the analysis. The Case order plot of Cook's distance for ECG and PPG signals are shown in Fig 2.

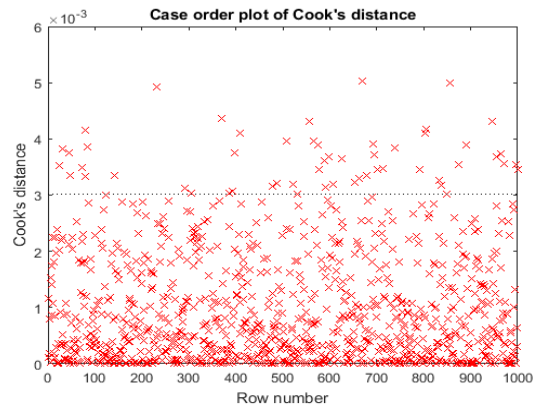


Fig (a)

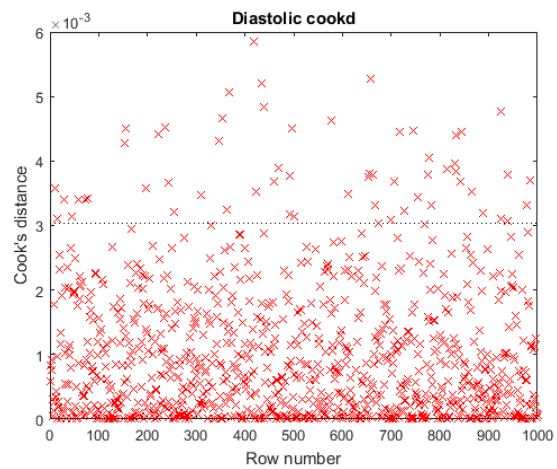


Fig (b)

Fig2. Cook's Distance plot for a) SBP and b) DBP

4.1.3 Normal Probability plot:

The Normal Probability Plot is a tool which is used for comparing the data set with normal distribution graphically. We can utilize it with the standardized residual of the linear regression model and check

whether the error term ϵ is distributed normally. The normal probability plot for both systolic data and diastolic data is shown in fig 3.

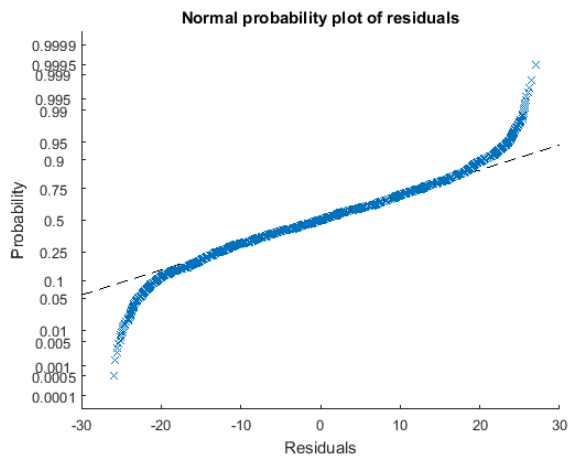


Fig (a)

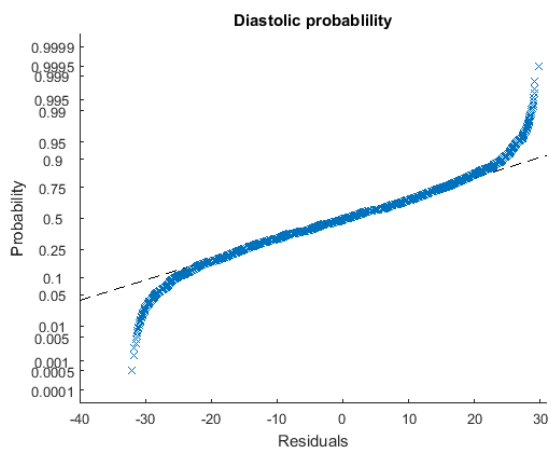


Fig (b)

Fig3. Normal probability plot for a) SBP and b) DBP

4.1.4 Lagged Residuals

Lagged residuals are used in regression analysis to give an approximation of effects for the independent variables. These residuals can be used to reduce the occurrence of autocorrelation in the model. The plot for residuals verses lagged residuals is shown in fig 4.

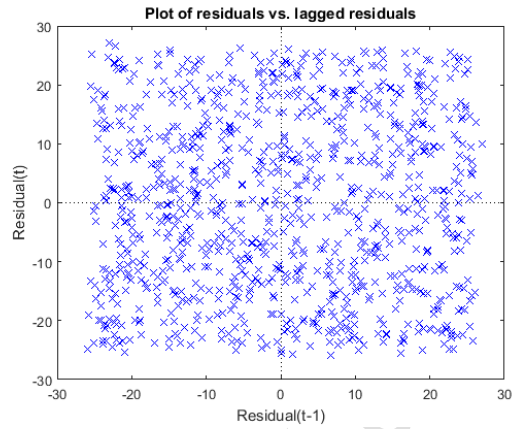


Fig (a)

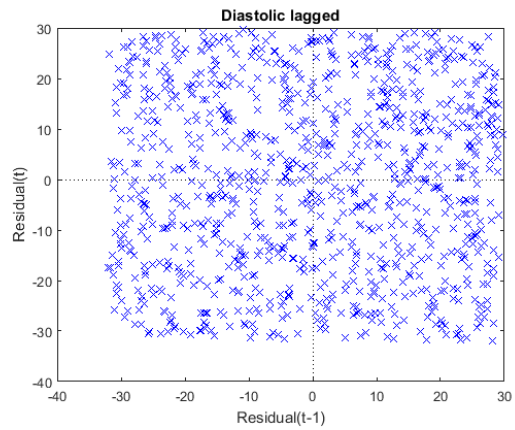


Fig (b)

Fig4.Residual vs Lagged residual plot for a) SBP and b) DBP

4.1.5 Fitted values

A Linear model, y = b0 + b1 x includes two constants, one is slope and other is intercept. The values of these constants are unknown, so they are called as unknown parameters of the model. To evaluate how well a specific linear model fits any of our data points(xi,yi) we should think about how well the model would anticipate the y-estimation of the point,

$$Yi = b0 + b1\ xi$$

These predictors in the dataset for the x values are called fitted values.

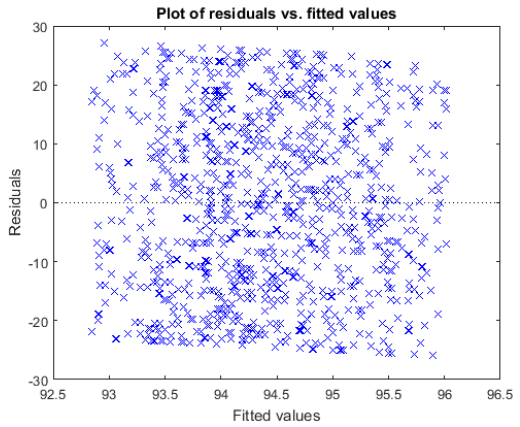


Fig (a)

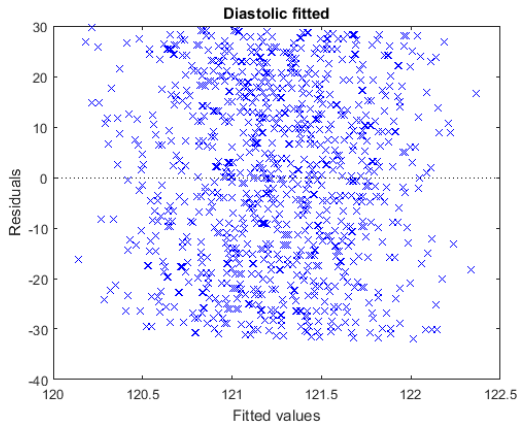


Fig (b)

Fig 5. Fitted value plot for a) SBP and b) DBP

These fitted values help to detect the nonlinearity of the model and the outliers. The plot for residuals verses fitted values for systolic and diastolic data are shown in Fig5.

4.1.6 Added variable plot

An Added variable plot which is also known as a partial regression plot shows you the correlation between the response variables and one of the predictors in the regression model, after managing the existence of the other predictors. Added variable plot for whole model is shown in fig 6.

IV. RESULT

The results for an efficient pre-processing and feature extraction methodology are shown below. The raw ECG and PPG signals which are used for the detection of BP are shown in fig 7. This signal contains the respiration artifacts and the base line wandering. For removal of these artifacts we use the db6 wavelet which is shown in fig 8. The glitches are removed from the signal by subtracting the original signals from wavelet. For the detection of the peak points from the signal a threshold value of 45 percent of the maximum signal is taken and the points are detected by using the threshold [8]. The filtered signal with detected peak points is shown in figure 9. The parameters are extracted from the ECG and PPG such as HRV analysis and the Poincare plot.

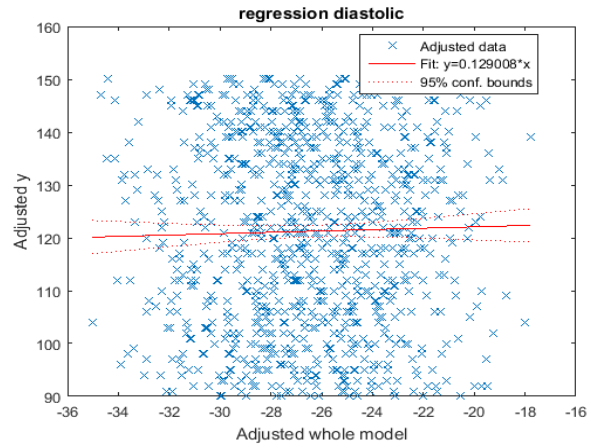


Fig (a)

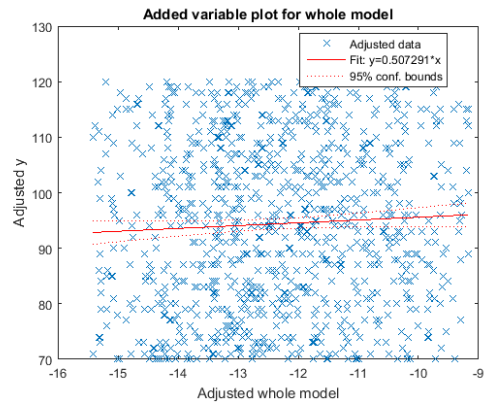


Fig (b)

Fig 6. Adjusted variable plot for a) SBP and b) DBP

Fig (a)

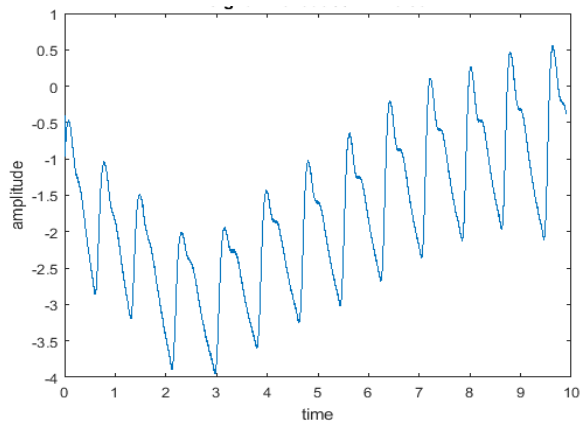


Fig (b)

Fig7. a) Raw ECG signal and b) Raw PPG signal.

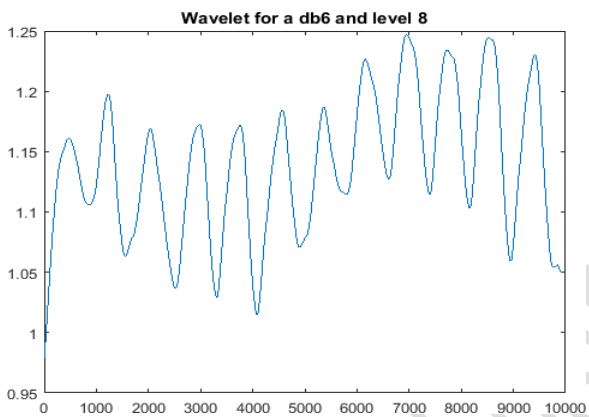
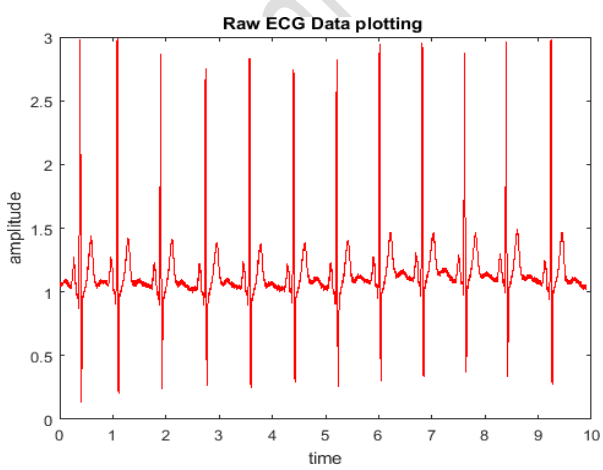


Fig (a)



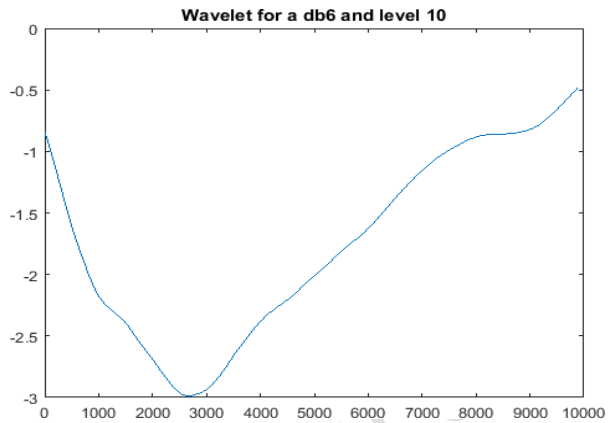


Fig (b)

Fig8. Daubechies wavelet for a) ECG signal and b) PPG signal.

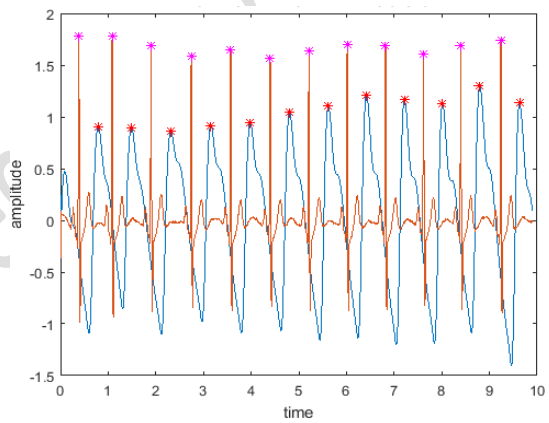


Fig9. ECG and PPG signal with peak points detected

During the Regression analysis the ANOVA model is performed on the Systolic and diastolic date for the regression analysis.

Table I:

Systolic information for regression

Input	Sum sq	DF	Mean sq	F	P value
X1	174.02	1	174.02	0.79242	0.37358
X2	403.79	1	403.79	1.8387	0.17541
X3	0.074366	1	0.074366	0.00033864	0.98532
erro	2.1873e+	99	219.6	-	-

r	05	6			

X new1 = (X1, X2, X3) = (-4.3816, 0.7986, 0.7978)

Table II:

Diastolic information for regression

Input	Sum sq	DF	Mean sq	F	P value
X1	30.618	1	30.618	0.097206	0.75527
X2	51.872	1	51.872	0.16468	0.68497
X3	95.979	1	95.979	0.30472	0.58106
error	3.1372e+05	996	314.98	-	-

X new2 = (-4.3816, 0.7986, 0.7978)

V. CONCLUSION

In this paper, the blood pressure is estimated using continuous and non invasive method. The proposed method contains preprocessing, Feature extraction and regression stages. We have shown that the proposed BP estimation algorithm works properly and the multiple regression analysis is used to obtain the values of blood pressure.

VI. REFERENCES

[1] W. H. Organization et al., World Health Statistic 2015. World Health Organization, 2015.
[2] L. Peter et al., "A review of methods for non-invasive and continuous blood pressure monitoring: Pulse transit time method is promising?" IRBM, vol. 35, no. 5, pp. 271–282, 2014.
[3] S. Ahmad et al., "Electrocardiogram-assisted blood pressure estimation," IEEE Trans. Biomed. Eng., vol. 59, no. 3, pp. 608–618, 2012.
[4] F. W. Xuan, "An exploration on real-time cuffless blood pressure estimation for e-home healthcare," Ph.D. dissertation, University of Macau, 2011.
[5] H. Gesche et al., "Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method," Eur. J. Appl. Physiol., vol. 112, no. 1, pp. 309–315, 2012.
[6] R. Mukkamala et al., "Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice," IEEE Trans. Biomed. Eng. vol. 62, no. 8, pp. 1879–1901, Aug 2015.
[7] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet : Components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215–e220, 2000.
[8] hossein robbani "R Peak Detection in Electrocardiogram Signal Based on an Optimal Combination of Wavelet Transform, Hilbert Transform, and Adaptive Thresholding"
[9] B. N. Singh et al., "Optimal selection of wavelet basis function applied to ECG signal denoising," Digital Signal Process., vol. 16, no. 3, pp. 275–287, 2006.