Pulse Pressure, Mean Arterial Pressure and Blood Pressure Prediction Based on PPG Signals

Bhavya Alok
Department Of Computer Science and Engineering
IIIT Naya Raipur
Naya Raipur, India
bhavya20100@iiitnr.edu.in

Kushal Jain
Department Of Computer Science and Engineering
IIIT Naya Raipur
Naya Raipur, India
kushal20100@iiitnr.edu.in

***Abstract*—The use of optical sensors is nowadays very common in the area of non-invasive diagnosis. This is mainly because of their salient features like, simple construction, low cost, easy to use, relatively inexpensive nature etc. Photoplethysmography (PPG) sensor is one among the wide variety of optical sensors available and is used to measure the blood volumetric changes occurring in the various parts of the body. PPG signal contains rich source of information related to cardio-pulmonary system. But the major problem associated with the signal is the motion artifacts, causing corruption in the original PPG signal. The aim of this paper is to use features extracted from a PPG signal to construct a machine learning model that can predict Pulse Pressure (PP), Mean Arterial Pressure (MAP), Diastolic Blood Pressure (DBP) and Systolic Blood Pressure (SBP).**

Keywords— Photoplethysmography, Pulse Pressure, Mean Arterial Pressure, Diastolic Blood Pressure, Systolic Blood Pressure, Machine Learning, Diagnosis

I. Introduction

Measurements of blood flow can be used to estimate blood volume variations in various areas of the body. Such blood volume measures are critical in clinical applications, as they can be used to diagnose a variety of biological illnesses such as arterial blockages, heart ailments, and so on. Plethysmographs are the instruments used to measure blood volume fluctuations, and Plethysmography is the technique. This method is based on the fact that blood absorbs more infrared light than the other tissues.

There are two parts to a PPG signal: an AC component and a DC component. The AC component of the PPG signal is obtained when light passes through arterial blood and is pulsatile. The absorption of light by blood in veins, bones, and tissues causes the DC component, or non-pulsatile portion. This signal comprises vital information such as heart rate variability, blood pressure, and breathing, among other things.

Many different types of PPG signals have been discovered, and they have been linked to ageing and cardiovascular disease. PPG signals are recorded from micro-vascular beds on the outside of the body, such as the finger, earlobe, forehead, and toe, in clinical practise. The PPG sensor's coverage region includes veins, arteries, and many capillaries. PPG waveforms have three different characteristics. A PPG waveform usually has a systolic peak, a diastolic peak, and a notch in between, as seen in Figure 1.

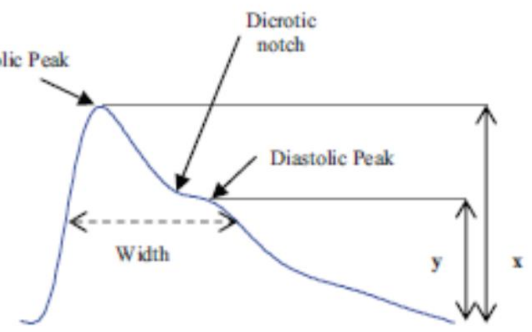


Fig. 1. A typical PPG Signal

PPG waveforms can be measured using two main methods: transmission and reflection. Light is emitted into the tissue, and a light detector is put on the opposite side of the tissue to measure the transmitted light in the transmission type. Because just a little amount of light travels through organ tissue, PPG is only useful for a few body parts, such as the finger or the ear lobe.

The light source and the light detector, on the other hand, are both located on the same side of a body part in the reflection type. The detector then measures the reflected light. This can be used on any region of the body because the reflected light is measured. The PPG signal from a commercially available transmission type pulse oximeter was used to process the data in this study.

The open-source dataset that was used in this work was analysed and evaluated by several research organisations. To assess the physiological changes in different levels of blood pressure, a novel technique for treating hypertension based on the theory of arterial wave propagation and the morphological theory of PPG was presented. To identify hypertension, ECG and PPG signals were acquired concurrently. An intrinsic association between the properties of systolic BP and PPG was developed using a model for PPG characteristics. Using the PPG pulse based on the pulse decomposition analysis, a PPG signal analysis was utilised to define obesity, age group, and hypertension.

T-domain, f-domain, (t, f)-domain, and statistical features are the most common features used for noninvasively determining blood pressure. Different groups employed a variety of t-domain features calculated from the original signal and its derivatives. In a previous study, the authors used frequency domain features to identify a neurological condition. In this study, the authors used Zaid et alwork .'s to generate features for properly calculating blood pressure from the PPG signal.

Several studies have documented various characteristics of the PPG signal for various purposes. These qualities have been employed by a variety of groups to assess SBP and DBP, although there is still room for improvement. As previously noted, a variety of automated ML approaches were evaluated and recorded for distinct PPG databases. To our knowledge, no recent research has used the machine learning techniques to incorporate t-, f-, and (t, f) domain characteristics to estimate BP with good accuracy. Because PPG signal processing is simpler and easier, emerging methods for extracting characteristics from PPG data are receiving more attention. To improve the accuracy of BP estimate using the PPG signal

II. MOTIVATION

Any illness that affects the heart or blood vessels is referred to as "cardiovascular disease." It's usually associated to atherosclerosis (fat deposits inside arteries) and an increased risk of blood clots.

Hypertension, often known as high blood pressure, is a chronic physiological condition that affects almost a billion people worldwide, and hypertensive individuals are more prone to develop additional cardiovascular problems. Although hypertension cannot be cured, it may be controlled via dietary adjustments, exercise, and other lifestyle changes [2].

Regular feedback on the success of such lifestyle changes is essential for creating a successful habit. Wearable technology is gaining popularity.

Using a sphygmomanometer to measure blood pressure is the gold standard for noninvasive blood pressure estimation. This technique, however, has several drawbacks, including the following:

1. A thick cuff that is difficult to travel with or wear for extended periods of time without being noticed.
2. Repeated blockage of the brachial artery owing to measurements, resulting in numbness and discomfort.
The amount of blood flowing through your arms may decrease.
3. Cuff inflation and deflation necessitate the employment of heavy electronics in combination with pneumatic systems (pump, valve, and battery).

Photo plethysmography, or PPG, is an indirect measure of vascular flow that has been shown to be strongly connected to changes in pressure wave or arterial blood pressure.

Blood pressure (BP) is a periodic signal that is proportional to heart rate in frequency. The upper bound of blood pressure is the Systolic Blood Pressure (SBP), while the bottom bound is the Diastolic Blood Pressure (DBP) (DBP). The mean arterial pressure is the average blood pressure during a cardiac cycle (MAP). Hypertension is defined as a blood pressure level of more than 140 millimeters of mercury (mmHg) or more than 90 millimeters of mercury (mmHg) that is potentially harmful to internal organs.

III. PROBLEM STATEMENT

To Predict Pulse Pressure, Blood Pressure and Mean Arterial Pressure using a dataset provided that contains:

a) 1.4 Million Singular Waveforms of length 126

b) Some extracted features from them like Ratio_B_A (calculated using doubly differentiating the signal)

c) Dependent variables (to predict)

To Create and learn to create a custom Decision tree regressor Model from Scratch.
To extract new features using PCA

IV. METHODOLOGY

In the start waveforms were plotted, plotting Waveforms for first few thousands of signals showed no significant difference.

All of them looked mostly like the below figure 2,

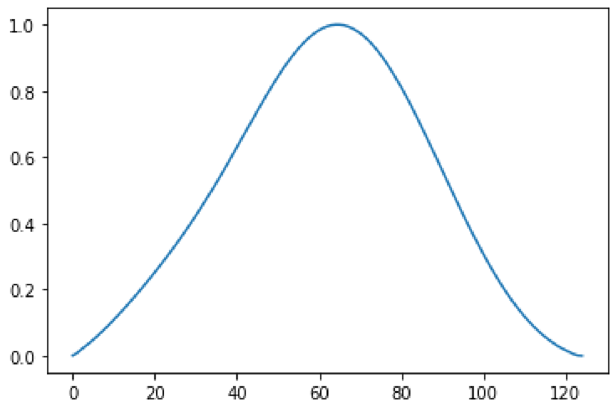


Fig 2: Sample Waveform from first few thousand Data points

Then to see the variance in the data plot of Dependent variable vs Entry number plotted which looked much like following:

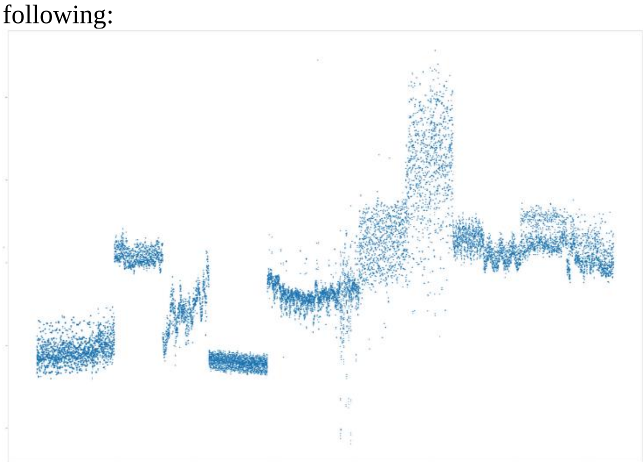


Fig 3: PP vs Data Point Plot

This Plot showed Existence of multiple classes and hence the reason the first few thousand plots were very much similar.

The waveforms of different classes were plotted individually, and as expected the waveforms of all different classes showed significant differences among each other

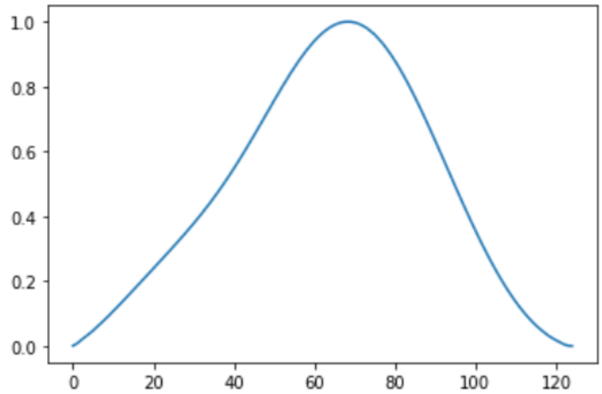


Fig 4: Waveforms of Class 1

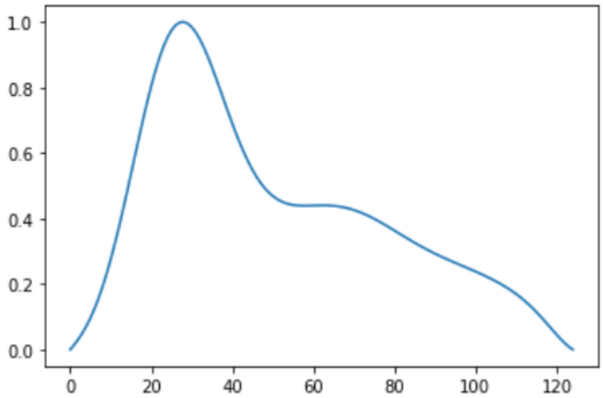


Fig 5: Waveforms of Class 2

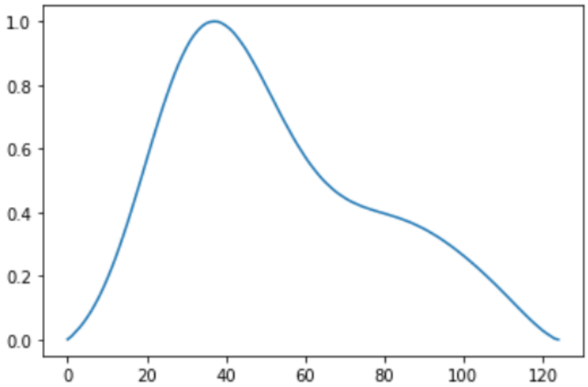


Fig 6: Waveforms of Class 3

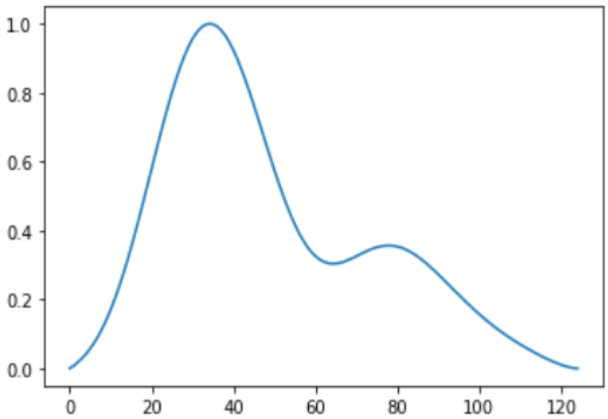


Fig 7: Waveforms of Class 4

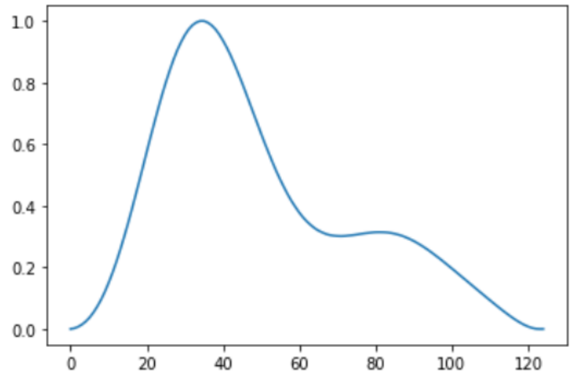


Fig 8: Waveforms of Class 5

To check the trend of features with respect to data, Features vs Entries plots were plotted.

Some features likely had no impact on the dependent variable this was obvious after the plots of dependent variable vs independent variables were plotted.

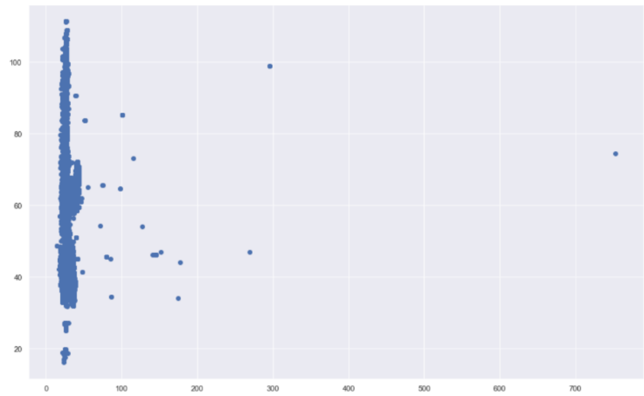


Fig 9: Crest vs PP, A plot of a useless feature

Others which are likely going to have significant impact on the dependent variable showed clusters in dependent variable vs independent variables plot.

The number of these clusters were similar in number to the number of clusters in the PP vs Data Points Plot.

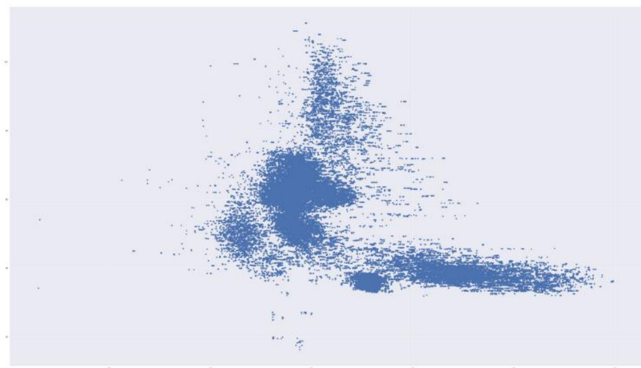


Fig 10: Ratio_E_A vs PP, A plot of potentially useful feature.

After Creating the correlation matrix, the following features were selected to build upon our models:

['PEAK_DISTANCE1','PEAK_DISTANCE2', 'DICROTIC', 'AUG_INDEX', 'S4', 'RATIO_E_A', 'DLASI1', 'DLASI2']

Standard Library Models of Random Forest and Decision Tree were used to get score of all the dependent variables.

The Results were as follows for Random Forest Regression:

	MSE	R^2
PP	1.13	0.9916
SBP	3.06	0.9892
DBP	1.04	0.9909
MAP	1.15	0.9914

The Results were as follows for Decision Tree Regression:

	MSE	R^2
PP	3.82	0.9805
SBP	6.82	0.9761
DBP	2.08	0.9818
MAP	2.42	0.9821

The Built-from-scratch Decision tree gave its results as follows (even after taking about 9.5 hours to execute, whereas the built-in models took no more than 30 Seconds to execute!)

Accuracy: 89.99%

2. Extraction of features using PCA was done after preprocessing.
The resultant was 8 new variables.

However, it seemed these only decreased the amount of information as the accuracies of all models went down.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8
0	-41.635774	17.726081	6.184566	1.677027	-1.240956	0.376384	-0.893928	0.014723
1	-40.365588	16.547268	7.709375	1.617388	-0.326955	-1.338357	0.666809	0.078733
2	-42.113276	16.270238	4.358151	-0.618525	-1.782528	1.339916	-2.604514	-1.903273
3	-43.196815	15.286254	5.600005	-0.524332	-0.577933	0.395595	-1.497244	-2.034589
4	-42.135820	16.997355	6.652072	2.059462	-0.128913	-0.691973	0.357410	0.044466
...	...	...	...	...	...	...	...	...
145822	0.202619	-0.241766	-3.939744	0.502834	1.522987	-1.541699	2.249898	-0.215124
145823	-0.371043	-2.664920	0.421976	3.132905	0.744836	-1.227483	2.554252	0.395826
145824	1.323824	-0.412215	0.516802	1.531047	1.083717	-3.174792	3.900460	-0.038085
145825	5.121064	0.921936	0.862627	-0.470325	2.067219	-0.570654	2.038141	-0.012547
145826	7.181939	-1.009004	1.051264	1.741622	0.280272	-1.481916	1.707919	-0.158968

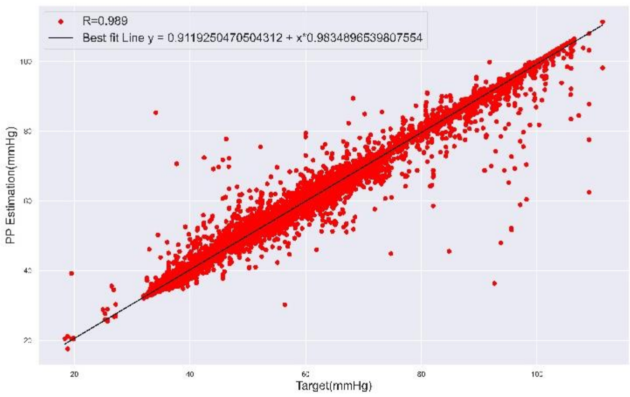
145827 rows × 8 columns

Fig 12: Features Extracted Using PCA

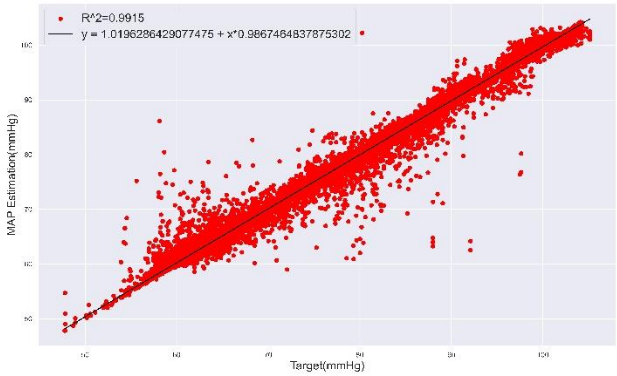
Accuracies before and after PCA,

Model, Target	Before PCA	After PCA
Random Forest Regressor, PP	99.16	97.64
Decision Tree Regressor, PP	98.05	94.57

Plot of y_test vs y_predict for PP,



Plot of y_test vs y_predict for MAP,



V. CONCLUSION

PPG signals have been reported as being able to be used to determine many cardiovascular parameters and arterial stiffness. In this paper, we see what PPG signals are and why they are used to detect so many cardiovascular parameters and arterial stiffness. There are different features that are extracted from PPG signals which have some relation with the target values Pulse Pressure (PP), Blood Pressure (BP) and Mean Arterial Pressure (MAP). Using these features, we find the features which vary highly with target values. Using different techniques of feature extraction and feature selection, we found some selected features which predicted all target values with high accuracy. We used two regression models to train the machine learning models which are Decision Tree and Random Forest. In the end, we could accurately predict the target values Pulse Pressure (PP), Blood Pressure (BP) and Mean Arterial Pressure (MAP).

VI. REFERENCES

[1] Chowdhury, Moajjem Hossain et al. "Estimating Blood Pressure from the Photoplethysmogram Signal and Demographic Features Using Machine Learning Techniques." *Sensors (Basel, Switzerland)* vol. 20,11 3127. 1 Jun. 2020, doi:10.3390/s20113127

[2] Kim, Seon-Chil & Cho, Sung-Hyoun. (2020). Blood Pressure Estimation Algorithm Based on Photoplethysmography Pulse Analyses. Applied Sciences. 10. 4068. 10.3390/app10124068.

[3] Zuhair, Aws. “A NOVEL WAVEFORM MIRRORING TECHNIQUE FOR SYSTOLIC BLOOD PRESSURE ESTIMATION FROM ANACROTIC PHOTOPLETHYSMOGRAM.” Journal of Engineering Science and Technology (2018): n. pag. Print. K. Elissa, “Title of paper if known,” unpublished.

[4] Bagha, Sangeeta & Shaw, Laxmi. (2011). A Real Time Analysis of PPG Signal for Measurement of SpO2 and Pulse Rate. INTERNATIONAL JOURNAL OF COMPUTER APPLICATIONS.

[5] G. Joseph, A. Joseph, G. Titus, R. M. Thomas and D. Jose, "Photoplethysmogram (PPG) signal analysis and wavelet de-noising," 2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD), 2014, pp. 1-5, doi: 10.1109/AICERA.2014.6908199.

[6] Elgendi, Mohamed. “On the analysis of fingertip photoplethysmogram signals.” *Current cardiology reviews* vol. 8,1 (2012): 14-25. doi:10.2174/157340312801215782.