ML_Project_Report

by Bhavya Alok

General metrics

13,312	2,016	186	8 min 3 sec	15 min 30 sec
characters	words	sentences	reading time	speaking time

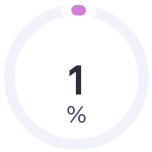
Score

This text scores better than 73% of all texts checked by Grammarly

Writing Issues

138	44	94
Issues left	Critical	Advanced

Plagiarism



2 sources

1% of your text matches 2 sources on the web or in archives of academic publications

Writing Issues

13	**Engagement**	
13	Word choice	
63	**Correctness**	
1	Text inconsistencies	
4	Comma misuse within clauses	
3	Confused words	
3	Misplaced words or phrases	
11	Determiner use (a/an/the/this, etc.)	
3	Mixed dialects of english	
12	Punctuation in compound/complex sentences	
6	Wrong or missing prepositions	
1	Incomplete sentences	
3	Incorrect noun number	
9	Misspelled words	
2	Conjunction use	
1	Closing punctuation	
1	Faulty subject-verb agreement	
1	Improper formatting	
2	Unknown words	
62	**Clarity**	
3	Intricate text	
11	Unclear sentences	
18	Wordy sentences	
29	Passive voice misuse	

1 Hard-to-read text

Unique Words

Measures vocabulary diversity by calculating the percentage of words used only once in your document

33%

unique words

Rare Words

Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

37%

rare words

Word Length

Measures average word length

5.1

characters per word

Sentence Length

Measures average sentence length

10.8

words per sentence

ML_Project_Report

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

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[1] in the area of non-invasive[2] diagnosis. This[3] is mainly because of their salient features like,[4] simple construction, low cost, easy[5] to use, relatively inexpensive nature etc. Photoplethysmography (PPG) sensor is one[7] among the wide variety of optical sensors available and is used to measure the blood volumetric[6] changes occurring in the[7] various parts of the body.[7] PPG signal contains rich[8] source of information related to cardio-pulmonary[9] system. But the major problem associated with the signal is the motion artifacts,[10] causing corruption in the

139

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

Introduction

Blood flow measurements can be used to estimate the blood volume changes in different parts of the body. Such blood volume measurements are of great importance in clinical applications which can be used to detect several biological disorders like arterial obstructions, cardiac diseases etc. Instruments used for measuring blood volume changes are called Plethysmographs and the technique is called Plethysmography. The idea behind this technique is that blood absorbs more infrared light than the remaining tissues.

A PPG signal consists of an AC component and a DC component. The pulsatile portion of the PPG signal is the AC component and is obtained when light passes through the arterial blood. The DC component or non-pulsatile portion is caused the absorption of light by blood in veins, bones and tissues. This signal contains important information about the heart rate variability, blood pressure, respiration etc.

Many different kinds of PPG signals have been identified and have been shown associated with age and cardiovascular pathology. In clinical practice, PPG signals are recorded from micro-vascular beds at exterior body locations, such as the finger, earlobe, forehead, and toe. The coverage area of the PPG sensor includes veins, arteries, and numerous capillaries. PPG waveforms generally

have three distinct features. As shown in Figure 1, a PPG waveform typically contains systolic peak, diastolic peak, and a notch in between.

Fig. 1. A typical PPG Signal

There are two different methods of measuring PPG waveforms: Transmission PPG and Reflection PPG. In transmission type, light is emitted into the tissue and a light detector is placed in the opposite side of the tissue to measure the transmitted light. As only a limited amount of light passes through the organ tissue the transmittance PPG is applicable only to a body parts such as the finger or the ear lobe.

However in the reflection type, the light source and the light detector are both placed on the same side of a body part. Then the reflected light is measured by the detector. Since the reflected light is measured, this can be used at any part of the body. In this work PPG signal obtained from a commercially available transmission type pulse oximeter was used for processing.

Several research groups have analyzed and evaluated the quality of the open-source dataset, which was used in this study. A novel approach for treating hypertension based on the theory of arterial wave propagation and morphological theory of PPG was proposed to check the physiological changes in different levels of blood pressure. ECG and PPG signals were obtained simultaneously to detect hypertension. A model for PPG characteristic was analyzed and an inherent relationship between the characteristics of systolic BP and PPG was established. In, a PPG signal analysis was used to characterize obesity, age group, and hypertension using the PPG pulse based on the pulse decomposition analysis.

The features typically used for noninvasively estimating BP are: (i) T-domain, (ii) f-domain, (iii) (t, f)-domain, (iv) and statistical features. Several t-domain features, which were calculated[53] from the original signal and its derivatives, were used by different groups. In a different[54] study,[55] showed the use of frequency domain[56] features for identifying a neurological disorder in[57] this study, the authors have taken inspiration from Zaid et al. to create features in estimating BP accurately from the PPG signal.

Several studies reported different features of the PPG signal for different[58] applications. Various groups have used these features for SBP and DBP measurement; however, there is still plenty of scope for improvement. Numerous[59] automated ML techniques were evaluated and recorded for various PPG databases as[60] mentioned earlier[59]. Nonetheless, to the best of[61] our knowledge, no recent work has combined t-, f-, and (t, f) domain features to estimate BP with a high accuracy using the machine learning approach. PPG signal processing is comparatively simpler[62] and easier,[63] so more attention is being[64] paid[65] to novel methods that extract features from PPG signals. To reduce the error in BP estimation based on the PPG signal,[66] this analysis not only extracts features from the PPG signal but also utilizes the demographic characteristics of subjects, such as height, weight, and[67,68] age, etc. There are several features that were[69] extracted[70,71] for BP estimation from the PPG signal in this study, which were not used before by any other group.

MOTIVATION

Any illness that affects the heart or blood vessels is referred[72] to as "cardiovascular disease." It's usually associated to[73] 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,[74] it may be controlled[75] 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[2] blood pressure estimation.[76] 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,[78] or PPG, is an indirect measure of vascular flow that has been shown[79] to be[80] strongly connected[81] to changes in pressure wave or arterial blood pressure.

Blood pressure (BP) is a periodic signal that is proportional to heart rate in frequency.[82] 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.

Problem Statement

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

1.4 Million Singular Waveforms of length 126

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

Dependent variables (to predict)

To Create and learn to create a custom Decision tree regressor Model from Scratch.

To extract new features using PCA

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

R2

PP

1.13

0.9916

SBP

3.06

0.9892

DBP

1.04

0.9909

MAP

1.15

0.9914

The Results[119] were as follows for Decision Tree Regression:

MSE[120]

R2

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[121]!)

Accuracy: 89.99%

Fig 11: Custom-Built Decision Tree R^2 Value

2. Extraction of features using PCA was done after preprocessing.
The resultant[122] was 8[123] new variables.

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

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

CONCLUSION

PPG signals have been reported[124] as being able[125] to be used to determine many cardiovascular parameters and arterial stiffness. In this paper, we see what PPG signals are[126] and why they are used to detect so many cardiovascular parameters and arterial stiffness. There are different features that are[127] extracted[128] 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[129] are Decision[130] Tree and Random Forest. In the end, we could accurately predict the target values Pulse[131] Pressure (PP), Blood Pressure (BP) and Mean Arterial Pressure (MAP).

References

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

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

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.

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

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.

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

1.	~~very common~~ → widespread	Word choice	Engagement
2.	*non-invasive; noninvasive*	Text inconsistencies	Correctness
3.	*This*	Intricate text	Clarity
4.	like,	Comma misuse within clauses	Correctness
5.	~~easy~~ → ease	Confused words	Correctness
6.	volumetric blood	Misplaced words or phrases	Correctness
7.	*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.*	Unclear sentences	Clarity
8.	a rich	Determiner use (a/an/the/this, etc.)	Correctness
9.	the cardio-pulmonary	Determiner use (a/an/the/this, etc.)	Correctness
10.	~~artifacts~~ → artefacts	Mixed dialects of English	Correctness
11.	This paper aims	Wordy sentences	Clarity
12.	*be used*	Passive voice misuse	Clarity
13.	, which	Punctuation in compound/complex sentences	Correctness
14.	*can be used*	Passive voice misuse	Clarity
15.	*Such blood volume measurements are of great importance in clinical applications which can be used to detect several biological disorders like arterial obstructions, cardiac diseases etc.*	Unclear sentences	Clarity

16.	, and	Punctuation in compound/complex sentences	Correctness
17.	*is obtained*	Passive voice misuse	Clarity
18.	by the	Wrong or missing prepositions	Correctness
19.	~~important~~ → essential	Word choice	Engagement
20.	*been shown*	Passive voice misuse	Clarity
21.	to be associated	Incomplete sentences	Correctness
22.	*are recorded*	Passive voice misuse	Clarity
23.	a systolic	Determiner use (a/an/the/this, etc.)	Correctness
24.	a diastolic	Determiner use (a/an/the/this, etc.)	Correctness
25.	*In transmission type, light is emitted into the tissue and a light detector is placed in the opposite side of the tissue to measure the transmitted light.*	Intricate text	Clarity
26.	*is emitted*	Passive voice misuse	Clarity
27.	, and	Punctuation in compound/complex sentences	Correctness
28.	~~in~~ → on	Wrong or missing prepositions	Correctness
29.	tissue,	Punctuation in compound/complex sentences	Correctness
30.	~~is applicable~~ → applies	Wordy sentences	Clarity
31.	~~a~~ body	Determiner use (a/an/the/this, etc.)	Correctness
32.	However,	Comma misuse within clauses	Correctness

33.	~~both~~	Wordy sentences	Clarity
34.	~~detector~~ → sensor	Word choice	Engagement
35.	*this can be used*	Passive voice misuse	Clarity
36.	~~part of the body~~ → body part	Wordy sentences	Clarity
37.	work,	Comma misuse within clauses	Correctness
38.	~~transmission type~~	Wordy sentences	Clarity
39.	dataset,	Punctuation in compound/complex sentences	Correctness
40.	*was used*	Passive voice misuse	Clarity
41.	*Several research groups have analyzed and evaluated the quality of the open-source dataset, which was used in this study.*	Unclear sentences	Clarity
42.	*was proposed*	Passive voice misuse	Clarity
43.	*A novel approach for treating hypertension based on the theory of arterial wave propagation and morphological theory of PPG was proposed to check the physiological changes in different levels of blood pressure.*	Unclear sentences	Clarity
44.	*ECG and PPG signals were obtained*	Passive voice misuse	Clarity
45.	~~characteristic~~ → characteristics	Incorrect noun number	Correctness
46.	*A model for PPG characteristic was analyzed*	Passive voice misuse	Clarity
47.	, and	Punctuation in compound/complex sentences	Correctness

48.	~~characteristics~~ → attributes	Word choice	Engagement
49.	*an inherent relationship between the characteristics of systolic BP and PPG was established*	Passive voice misuse	Clarity
50.	*A model for PPG characteristic was analyzed and an inherent relationship between the characteristics of systolic BP and PPG was established.*	Unclear sentences	Clarity
51.	~~In, a~~ → A	Wordy sentences	Clarity
52.	*was used*	Passive voice misuse	Clarity
53.	*were calculated*	Passive voice misuse	Clarity
54.	~~a different~~ → a separate, a further, an additional, an other	Word choice	Engagement
55.	study,	Punctuation in compound/complex sentences	Correctness
56.	frequency-domain	Misspelled words	Correctness
57.	~~in~~ → . In, ; in	Punctuation in compound/complex sentences	Correctness
58.	~~different~~ → other	Word choice	Engagement
59.	*Numerous automated ML techniques were evaluated and recorded for various PPG databases as mentioned earlier.*	Unclear sentences	Clarity
60.	, as	Punctuation in compound/complex sentences	Correctness
61.	~~the best of~~	Wordy sentences	Clarity
62.	~~simpler~~ → more straightforward	Word choice	Engagement

63.	~~easier~~ → more accessible, more manageable	Word choice	Engagement
64.	~~being~~	Wordy sentences	Clarity
65.	*being paid*	Passive voice misuse	Clarity
66.	*To reduce the error in BP estimation based on the PPG signal*	Misplaced words or phrases	Correctness
67.	~~and~~	Conjunction use	Correctness
68.	~~and~~	Wordy sentences	Clarity
69.	Several features were	Wordy sentences	Clarity
70.	*were extracted*	Passive voice misuse	Clarity
71.	*were extracted*	Passive voice misuse	Clarity
72.	*is referred*	Passive voice misuse	Clarity
73.	~~to~~ → with	Wrong or missing prepositions	Correctness
74.	*hypertension cannot be cured*	Passive voice misuse	Clarity
75.	*it may be controlled*	Passive voice misuse	Clarity
76.	*Using a sphygmomanometer to measure blood pressure is the gold standard for noninvasive blood pressure estimation.*	Hard-to-read text	Clarity
77.	~~periods of time~~ → periods	Wordy sentences	Clarity
78.	Photoplethysmography	Confused words	Correctness
79.	*been shown*	Passive voice misuse	Clarity
80.	~~has been shown to be~~ → is	Wordy sentences	Clarity

81.	*be strongly connected*	Passive voice misuse	Clarity
82.	*Blood pressure (BP) is a periodic signal that is proportional to heart rate in frequency.*	Unclear sentences	Clarity
83.	*is defined*	Passive voice misuse	Clarity
84.	~~defined as~~	Wordy sentences	Clarity
85.	~~millimeters~~ → millimetres	Mixed dialects of English	Correctness
86.	~~millimeters~~ → millimetres	Mixed dialects of English	Correctness
87.	*custom Decision tree regressor Model*	Intricate text	Clarity
88.	~~In~~ → At	Wrong or missing prepositions	Correctness
89.	start,	Comma misuse within clauses	Correctness
90.	and plotting	Conjunction use	Correctness
91.	the first	Determiner use (a/an/the/this, etc.)	Correctness
92.	~~looked mostly~~ → mainly looked, looked mainly	Word choice	Engagement
93.	the first	Determiner use (a/an/the/this, etc.)	Correctness
94.	the Existence	Determiner use (a/an/the/this, etc.)	Correctness
95.	*This Plot showed Existence of multiple classes and hence the reason the first few thousand plots were very much similar.*	Unclear sentences	Clarity
96.	*were plotted*	Passive voice misuse	Clarity
97.	, the	Punctuation in	Correctness

		compound/complex sentences	
98.	~~classes~~ → types, categories, styles	Word choice	Engagement
99.	~~among~~ → from	Wrong or missing prepositions	Correctness
100.	other.	Closing punctuation	Correctness
101.	~~with respect to~~ → concerning, for, to	Wordy sentences	Clarity
102.	*To check the trend of features with respect to data*	Misplaced words or phrases	Correctness
103.	*Features vs Entries plots were plotted*	Passive voice misuse	Clarity
104.	~~this was~~ → . This was, ; this was	Punctuation in compound/complex sentences	Correctness
105.	*the plots of dependent variable vs independent variables were plotted*	Passive voice misuse	Clarity
106.	a significant	Determiner use (a/an/the/this, etc.)	Correctness
107.	the dependent	Determiner use (a/an/the/this, etc.)	Correctness
108.	*Others which are likely going to have significant impact on the dependent variable showed clusters in dependent variable vs independent variables plot.*	Unclear sentences	Clarity
109.	~~The number of these~~ → These	Wordy sentences	Clarity
110.	~~were~~ → was	Faulty subject-verb agreement	Correctness
111.	~~clusters~~ → groups, sets, collections	Word choice	Engagement
112.	~~useful~~ → helpful	Word choice	Engagement

113.	~~feature~~ → features	Incorrect noun number	Correctness
114.	*the following features were selected*	Passive voice misuse	Clarity
115.	a score, or the score	Determiner use (a/an/the/this, etc.)	Correctness
116.	~~score~~ → scores	Incorrect noun number	Correctness
117.	~~Results~~ → results	Misspelled words	Correctness
118.	~~MSE~~ → MAP, MOUSE, ME	Misspelled words	Correctness
119.	~~Results~~ → results	Misspelled words	Correctness
120.	~~MSE~~ → MAP, MOUSE, ME	Misspelled words	Correctness
121.	~~execute~~ → complete, run, achieve, accomplish	Word choice	Engagement
122.	~~resultant~~ → result	Confused words	Correctness
123.	~~8~~ → eight	Improper formatting	Correctness
124.	*been reported*	Passive voice misuse	Clarity
125.	~~as being able~~	Wordy sentences	Clarity
126.	PPG signals	Wordy sentences	Clarity
127.	Different features are	Wordy sentences	Clarity
128.	*are extracted*	Passive voice misuse	Clarity
129.	, which	Punctuation in compound/complex sentences	Correctness
130.	*We used two regression models to train the machine learning models which are Decision Tree and Random Forest.*	Unclear sentences	Clarity

131.	of Pulse	Wrong or missing prepositions	Correctness
132.	*PHOTOPLETHYSMOGRAM*	Unknown words	Correctness
133.	~~Real Time~~ → Real-Time	Misspelled words	Correctness
134.	*A Real Time Analysis of PPG Signal for Measurement of SpO2 and Pulse Rate.*	Unclear sentences	Clarity
135.	~~iCMMD~~ → MD	Misspelled words	Correctness
136.	~~doi~~ → DOI	Misspelled words	Correctness
137.	~~Elgendi~~ → Elgen	Misspelled words	Correctness
138.	*photoplethysmogram*	Unknown words	Correctness
139.	*The aim of this paper is to use*	A Model-based Systems Engineering Approach To Trade Space Exploration Of Virtual Reality Training Systems	Originality
140.	*A Real Time Analysis of PPG Signal for Measurement of SpO2 and Pulse Rate. INTERNATIONAL JOURNAL OF COMPUTER APPLICATIONS.*	IJCA - A Real Time Analysis of PPG Signal for Measurement of SpO2 and ... https://www.ijcaonline.org/archives/volume36/number11/4537-6461	Originality