Improved Discrimination Between Healthy and Hypertensive Individuals Combining Photoplethysmography and Electrocardiography

Jesús Cano[1], Fernando Hornero[2], Aurelio Quesada[3],
Arturo Martínez-Rodrigo[4], Raúl Alcaraz[4], José J Rieta[1*]

[1] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain
[2] Cardiovascular Surgery Department, Hospital Clínico Universitario de Valencia, Spain
[3] Cardiology Department, Hospital General Universitario de Valencia, Spain
[4] Research Group in Electronic, Biomed. and Telecomm. Eng., Univ. of Castilla-La Mancha, Spain

Abstract

Cardiovascular disease is one of the leading causes of death, with hypertension (HT) being its main risk factor. Its complications can be avoided with early treatment, but since these patients do not present any symptoms, HT is often detected at very advanced stages. This work presents a model for estimating blood pressure (BP) from electrocardiographic (ECG) and photoplethysmographic (PPG) signals, which can be easily obtained by means of wearable continuous monitoring devices. ECG, PPG and BP recordings from 86 patients were analyzed.A total of 34 standard and new features based on previous works were defined, such as pulse arrival times (PAT), and morphological characteristics of PPG signal. 37 classification models, ranging from Logistic Regression, Support Vector Machines (SVM), Nearest Neighbors, Naive Bayes or Coarse Trees were trained to compare discrimination results. The classifier that provided the highest performance when comparing normotensive patients with prehypertensive and hypertensive patients were Coarse Tree, providing an F1 score of 85.44% *(Se of* 86.27% *and Sp of* 77.14%*). The use of PPG and ECG features has successfully discriminated between healthy and hypertensive individuals and, thus, could be used to detect HT by embedding these techniques in wearable devices.*

1. Introduction

High blood pressure (HBP) is a major risk factor for many cardiovascular diseases worldwide, including heart disease, vascular diseases of the brain, and blood vessels diseases, specially in high-income countries [1]. Hypertension (HT) is a condition in which the blood vessels have persistently raised pressure. Regular blood pressure (BP) control is crucial for patients already suffering from HT, as many of their organs are particularly vulnerable to elevated BP. Healthy lifestyles, early detection and regular assessment of blood pressure levels, together with the obtention of a proper diagnosis are all beneficial for the prevention and control of HT and its consequences [2].

For noninvasive BP estimation, conventional cuff-based BP measurement devices offer good accuracy. However, they are not designed to be wearable, are not compatible with continuous measurement throughout the day, are uncomfortable, and their measurement procedure is somewhat tedious and requires patient attention [3]. To this must be added that most patients with HT have no symptoms in the elevated blood pressure stage and even in hypertension, therefore, many people miss, through ignorance and lack of medical control, the opportunity for early treatment and experience cardiovascular complications that could be avoided [1].

Because of the above factors, work in this field is focused on the development of robust and discrete BP estimation systems that can provide the user with regular updates of the BP level in near real time [4]. The main advance that has led to the development of these systems is the increasing presence of wearable devices, such as wristbands or smart watches [5], capable of monitoring physiological signals that change as a function of BP level, such as the electrocardiogram (ECG) and photoplethysmogram (PPG). Morphological changes in physiological signals mainly reflect changes in the state of functioning of the heart and vascular system, so morphological information from PPG could be used to assess hypertension [6].

The present work uses the analysis and processing of ECG and PPG signals to develop an improved system for discriminating between normotensive, prehypertensive and hypertensive patients. The final goal is to identify, by means of wearable devices, hypertensive patients without apparent symptoms or to monitor at risk patients in order to prevent future cardiovascular diseases, of which HT is the main risk factor.

 ISSN: 2325-887X DOI: 10.22489/CinC.2021.030

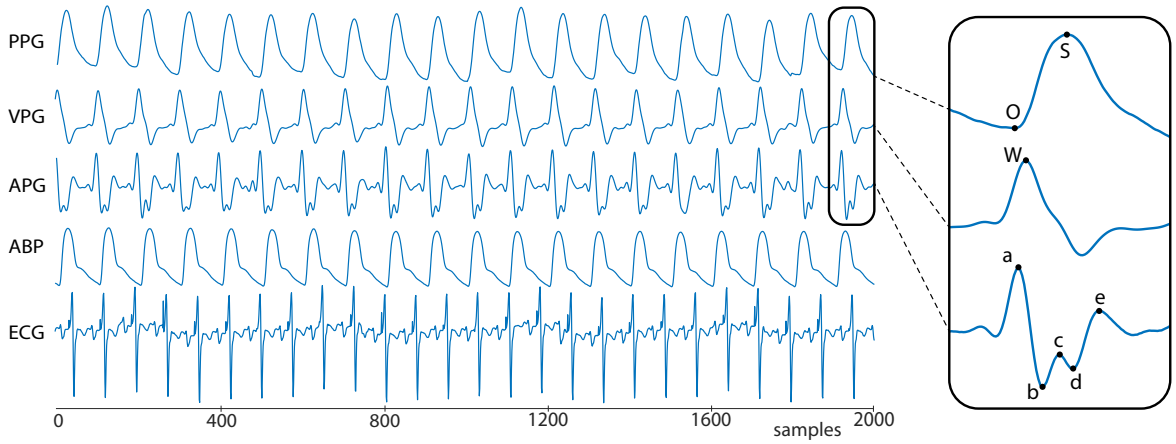

Figure 1. Fragment of 2000 samples illustrating the morphology of the signals used, together with the representation of the characteristic points of the defined PPG, VPG and APG signals.

2. Material and Methods

In this study, the signals used were obtained from MIMIC database, a free database that contains information from Intensive Care Unit (ICU) patients admitted to Beth Israel Deaconess Medical Center in Boston, USA [7].

A total of 86 patients records distributed in 35 normotensive, 26 prehypertensive and 25 hypertensive containing ECG, PPG and BP (ABP) signals recorded simultaneously and in acceptable conditions were analysed. The records with noise, abnormal morphology or missing data were excluded. The signals were all recorded simultaneously with a duration of 60 seconds, a common sampling frequency of 125 Hz and a resolution of 8-10 bits [8].

2.1. Data Preprocessing and feature points

Firstly, in order to remove noise and improve signal quality, a fourth order Chebyshev II bandpass filter was applied to the raw PPG signal between 0.5 and 10 Hz [9]. Baseline fluctuation was then removed by setting the pulse minimum to zero from the difference between the signal and its lower envelope to improve the precision obtaining the amplitudes of the PPG feature points.

From the PPG signal, its first and second derivatives, called photoplethysmographic velocity signal (VPG) and photoplethysmographic acceleration signal (APG), respectively, were obtained [10]. At last, their characteristic points have been acquired, searching for local maximum and minimum calculated based on thresholds in each of the pulses of the signals, as illustrated in Figure 1. The systolic peaks of the three signals (S,W,a), the diastolic notches of the PPG signal (O), and two local maxima and minimum of the APG signal (b,c,d,e) were extracted [10].

ABP signals were clear and did not require any processing to be applied. Only the absolute maximum systolic blood pressure (SBP) had to be detected, used as the BP classification label. Finally, standard preprocessing was applied to each ECG [11]. Thus, it was high-pass filtered with cutoff frequency of 0.5 Hz to remove the baseline, and then low-pass filtered with cutoff frequency of 50 Hz to reduce high-frequency muscle noise and remove power grid interference, in this case, 60 Hz [11]. An R-peak detector was then applied to obtain the position of each beat [12].

2.2. Obtaining discriminatory parameters

Two data matrices were used as inputs for the classification models. The first matrix was formed with the 13 parameters that reported a higher correlation with BP levels in previous works [13], such as the three pulse arrival times (PAT) that were calculated as the delay between the R peak of the ECG and a characteristic point of the PPG signal, PAT_{peak}, $\text{PAT}_{derivate}$, PAT_{foot}, time intervals between peak S from PPG signal and points c and d from APG signal (see Fig. 1), slopes between feature points, ratios between amplitudes and areas under the pulse.

The other matrix was formed with 24 parameters proposed used in the literature based on the three PATs and PPG signal characteristics, such as areas before and after the systolic peak (A1 and A2), the ratio between them, the total area, intervals between systolic peaks (TPP), total pulse interval (TPI), rise time and PTT, interval between the SBP peak and the systolic peak, amplitude of the systolic peaks in the PPG and VPG signals and ratios between the characteristic points of the APG signal with the aim of

Normotensive vs Prehypertensive	Normotensive vs Hipertenso	Norm. + Prehyp. vs Hypertensive	Normotensive vs Preyp. + Hip.

Table 1. Types of patients compared in each matrix

	AdaBoost			Logistic Regresion			KNN			Bagged		
	Se	Sp	F_1	Se	Sp	F_1	Se	Sp	F_1	Se	Sp	F_1
Normo vs Pre	61,54%	74,29%	62,75%	57,69%	68,57%	57,69%	42,31%	80,00%	50,00%	46,15%	74,29%	51,06%
Normo vs Hiper	80,00%	74,29%	74,07%	52,00%	77,14%	56,52%	40,00%	94,29%	54,05%	76,00%	82,86%	76,00%
Normo+Pre vs Hiper	88,00%	81,97%	75,86%	36,00%	90,16%	45,00%	36,00%	100,00%	52,94%	52,00%	90,16%	59,09%
Normo vs Pre+Hiper	76,47%	71,43%	78,00%	60,78%	34,29%	59,05%	62,75%	68,57%	68,09%	72,55%	57,14%	71,84%

Table 2. Performance of the four classification models analysed with the characteristic parameters that have been reported to correlate most strongly with BP levels in previous work.

	Naive Bayes			SVM cubic			SVM quadratic			Coarse Tree		
	Se	Sp	F_1	Se	Sp	F_1	Se	Sp	F_1	Se	Sp	F_1
Normo vs Pre	57,69%	91,43%	68,18%	57,69%	74,29%	60,00%	61,54%	91,43%	71,11%	50,00%	80,00%	56,52%
Normo vs Hiper	64,00%	85,71%	69,57%	68,00%	82,86%	70,83%	60,00%	88,57%	68,18%	40,00%	80,00%	47,62%
Normo+Pre vs Hiper	64,00%	91,80%	69,57%	64,00%	91,80%	69,57%	68,00%	93,44%	73,91%	48,00%	77,05%	47,06%
Normo vs Pre+Hiper	82,35%	82,86%	84,85%	54,90%	65,71%	61,54%	76,47%	80,00%	80,41%	86,27%	77,14%	85,44%

Table 3. Performance of the four new proposed classification models using the new characteristic parameters.

improving the results of previous work [10].

Since three different groups of patients were available, the four most logical paired comparisons were made and shown in the Table 1. Furthermore, a treatment of outliers has been carried out, where the detected values are replaced by the median of the patients sharing their same classification label following the absolute deviation from the median (MAD) method [14].

2.3. Classification models

Initially, Logistic Regression, AdaBoost Tree, K Nearest Neighbors (KNN) and Bagged Tree models were applied, with representative classification theories such as regression, decision trees, cluster and bagged decision tree [15], since they have reported good results in previous works [13].

Furthermore, four additional classification models have also been applied searching for an improve in the accuracy percentaje. For this purpose, up to 37 different classification strategies have been tested, such as various types of decision trees, discriminant analysis, logistic regression, Naive Bayes, support vector machines (SVM), KNN and ensemble classifiers [15]. Finally, the cubic and quadratic SVM models, Naive Bayes using kernels and Coarse Tree were selected for obtaining the highest percentages of classificatory accuracy.

2.4. Statistical analysis

The cross-validation leaving one out strategy was used to estimate the models accuracy. Finally, statistical tests have been used to perform an evaluation of the classification. In this work, sensitivity (Se) or ability to detect the disease in diseased subjects, specificity (Sp) or ability to give as negative healthy patients and F1 score, which is the harmonic mean of accuracy of detecting false positives and sensitivity are used.

3. Results

The statistical study began using the predictive parameters that have reported a higher correlation with BP levels in previous work [13] and following the four groupings of patients in Table 1 as represented in Table 2. It can be seen that the best classifications are obtained by comparing normotensive patients with prehypertensive and hypertensive patients in the AdaBoost model with an F1 score of 78%. The other models do not discriminate correctly and tend to classify as healthy both patients who are really healthy and those who are sick as seen in the imbalance between Se and Sp generalized in these models.

Similarly, Table 3 presents the improved results of this study, where the use of the new predictive parameters, together with the alternative classification models, significantly improve the classification results. Thus, the value of F1 score for the Naive Bayes and Coarse Tree classification models exceeds 84%.

4. Discussion

The study of the variation in the morphology of the PPG signal has a large amount of physiological information that can be used to study cardiovascular activity. There are works in which only the PAT is used [6, 16], but whose effectiveness has been discussed in later works by combining it with additional PPG characteristics, which allowed reporting a higher correlation with BP levels [13]. Thus,

PAT indicates the transmission of the arterial wave in the blood vessel, whereas PPG features indicate the change of state of the vascular tissue and blood volume.

So far, recent work that has used the PPG signal both to classify patients as healthy or hypertensive and to obtain the level of BP has not agreed on the predictive parameters or the classification models to be used, since this choice depends largely on the signals used in the studies and there is disparity in terms of patient selection, mode of acquisition, and signal quality.

It is noteworthy that, in all the models analyzed in the present work, the best results are obtained when normotensive patients are compared with prehypertensive and hypertensive patients. This fact is very relevant, since it means that the values of the discriminatory parameters in prehypertensive patients are more similar to hypertensive patients. Furthermore, it should be taken into account that prehypertensive patients do not show obvious symptoms until they are in very advanced stages of the disease, causing serious cardiovascular problems, so that alerting this group as sick patients is of great interest.

Finally, this study has certain limitations. The number of patients is not too large, there is no information on factors that may imply a higher risk of hypertension such as physical condition, sex or age, and it would be desirable to use signals with a sampling frequency higher than 125Hz, which could improve the identification of fiducial points.

5. Conclusion

This work has shown that the combined analysis of PPG and ECG signals, together with the definition of new morphological characteristics, such as pulse width and interval between systolic peaks, as well as the use of alternative classification models, allows better discrimination between healthy individuals and prehypertensive or hypertensive patients. The implementation of these methodologies in wearable devices may enable the prevention of hypertension and its associated cardiovascular pathologies.

Acknowledgements

Research supported by grants DPI2017–83952–C3 from MINECO/AEI/FEDER UE, SBPLY/17/180501/000411 from JCCLM and AICO/2021/286 from GVA.

References

[1] Unger T, Borghi C, Charchar F, et al. 2020 International Society of Hypertension Global Hypertension Practice Guidelines. Hypertension jun 2020;75(6):1334–1357.

[2] Levy D. The Progression From Hypertension to Congestive Heart Failure. JAMA The Journal of the American Medical Association may 1996;275(20):1557. ISSN 0098-7484.

[3] Frese EM, Fick A, Sadowsky SH. Blood Pressure Measurement Guidelines for Physical Therapists. Cardiopulmonary Physical Therapy Journal jun 2011;22(2):5–12.

[4] Kario K. Management of Hypertension in the Digital Era. Hypertension sep 2020;76(3):640–650. ISSN 0194-911X.

[5] Kulkarni S. Hypertension Management in 2030: a Kaleidoscopic View. Jrnl of Human Hypertension nov 2020;1–6.

[6] Mukkamala R, Hahn JO, Inan OT, Mestha LK, Kim CS, Toreyin H, Kyal S. Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice. IEEE Trans Biomed Eng aug 2015;62(8):1879–1901.

[7] Johnson AE, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG. MIMIC-III, a Freely Accessible Critical Care Database. Scientific Data may 2016;3(1):1–9.

[8] Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation jun 2000;101(23).

[9] Liang Y, Elgendi M, Chen Z, Ward R. Analysis: An Optimal Filter for Short Photoplethysmogram Signals. Scientific Data may 2018;5(1):1–12. ISSN 20524463.

[10] Elgendi M. On the Analysis of Fingertip Photoplethysmogram Signals. Current Cardiology Reviews jun 2012; 8(1):14–25. ISSN 1573403X.

[11] Sörnmo L, Laguna P. Bioelectrical Signal Processing in Cardiac and Neurological Applications. Elsevier, 2005. ISBN 9780124375529.

[12] Martínez A, Alcaraz R, Rieta JJ. Application of the Phasor Transform for Automatic Delineation of Single-Lead ECG Fiducial Points. Physiological Measurement nov 2010; 31(11):1467–1485. ISSN 09673334.

[13] Liang Y, Chen Z, Ward R, Elgendi M. Hypertension Assessment via ECG and PPG Signals: An Evaluation Using MIMIC Database. Diagnostics sep 2018;8(3):65.

[14] Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting Outliers: Do not use Standard Deviation around the Mean, use Absolute Deviation around the Median. Journal of Experimental Social Psychology jul 2013;49(4):764–766.

[15] Shalev-Shwartz S, Ben-David S. Understanding Machine Learning: From Theory to Algorithms, volume 9781107057. Cambridge University Press, jan 2013. ISBN 9781107298019.

[16] Liang Y, Abbott D, Howard N, Lim K, Ward R, Elgendi M. How Effective Is Pulse Arrival Time for Evaluating Blood Pressure? Challenges and Recommendations from a Study Using the MIMIC Database. Journal of Clinical Medicine mar 2019;8(3):337. ISSN 2077-0383.

Address for correspondence:

José J. Rieta

BioMIT.org, Electronic Engineering Department, Building 7F-5

Universitat Politecnica de Valencia

Camino de Vera, s/n, 46022, Valencia, Spain

E-mail: jjrieta@upv.es