Novel Photoplethysmographic and Electrocardiographic Features for Enhanced Detection of Hypertensive Individuals

Jesús Cano[1], Aurelio Quesada[2], Flavia Ravelli[3], Roberto Zangróniz[4], Raúl Alcaraz[4], José J. Rieta[1]*

[1] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain; {jecaser, jjrieta}@upv.es

[2] Cardiology Department, Hospital General Universitario de Valencia, Valencia, Spain; quesada_aur@gva.es

[3] Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Italy; flavia.ravelli@unitn.it

[4] Research Group in Electronic, Biomedical and Telecomm. Eng., Univ. of Castilla-La Mancha, Spain; raul.alcaraz@uclm.es

***Abstract*—Hypertension is a major risk factor for many cardiovascular diseases, which are the leading cause of death worldwide. Regular monitoring is essential to provide early diagnosis since most patients with elevated blood pressure (BP) have asymptomatic hypertension. This work presents a method for BP classification using simultaneous electrocardiographic (ECG), photoplethysmographic (PPG) and BP signals. 86 recordings were used, being 35 normotensive, 26 prehypertensive and 25 hypertensive. It has been proposed 23 novel features to improve the discrimination between healthy and hypertensive individuals based on pulse arrival times (PAT) and morphological features from PPG, VPG and APG signal. Moreover, alternative classification models as Support Vector Machines (SVM), Naive Bayes or Coarse Trees were trained with the defined features to compare the classification performance. The classifier that provided the highest results comparing normotensive with prehypertensive and hypertensive subjects were Coarse Tree, providing an F1 score of 85.44% (Se of 86.27% and Sp of 77.14%). The use of new PPG and ECG features has successfully improved the discrimination between healthy and hypertensive individuals, around 7% of F1 score, so this machine learning methodology would be of high interest to detect HT introducing these techniques in wearable devices.**

***Keywords*—Photoplethysmogram (PPG); blood pressure (BP); machine learning (ML); clasiffication models**

I. Introduction

Many cardiovascular diseases, including heart failure, blood vessels diseases, and vascular diseases of the brain are mainly caused by elevated blood pressure (BP) or hypertension [1]. Regular BP control is crucial for people already suffering from hypertension (HT), as they are particularly vulnerable to any physiological situation where elevated BP can befall. Early detection and treatment of hypertensive subjects, clinically controlled regular assessment of blood pressure levels and healthy lifestyles, combined with accurate diagnosis, are all beneficial for the control and prevention of HT [2].

Conventional BP measurement devices that get use of a cuff applied on the brachial artery for noninvasive BP estimation offer elevate accuracy and are extremely extended both in clinical and home settings. However, they are not wearable and compatible with continuous measurement during the day, are annoying for the subject, and their measurement proceeding requires patient attention and professional knowledge [3]. Moreover, most patients with HT are asymptomatic or have slight symptoms when subjects present elevated blood pressure levers and even in hypertension. Therefore, many people miss, through lack of medical control and preventive methods, the opportunity for early hypertension therapy and experience cardiovascular problems that might be avoided [1].

As a result of these problems in BP monitoring, recent studies are focused on the development of noninvasive and robust BP estimation methods that can provide the user with periodic information of the BP level throughout the day [4]. Wearable devices, such as smart watches have facilitated the increasing development of new methods for BP monitoring [5]. This devices use sensors to monitor physiological signals such as the electrocardiogram (ECG) and photoplethysmogram (PPG) that change as a function of BP level. Changes in the functioning of the heart and vascular system are mainly reflected as morphological changes in physiological signals, so morphological information obtained from PPG signal variation could be used for hypertension assessment [6].

The aim of this study is to propose novel photoplethysmographic and electrocardiographic morphological features to develop an improved system to discriminate between normotensive, prehypertensive and hypertensive patients. To prove the discriminatory power of these features, machine learning based classification models will be trained to identify hypertensive patients without apparent symptoms or to monitor continuously at risk patients for the prevention and early diagnosis of hypertension.

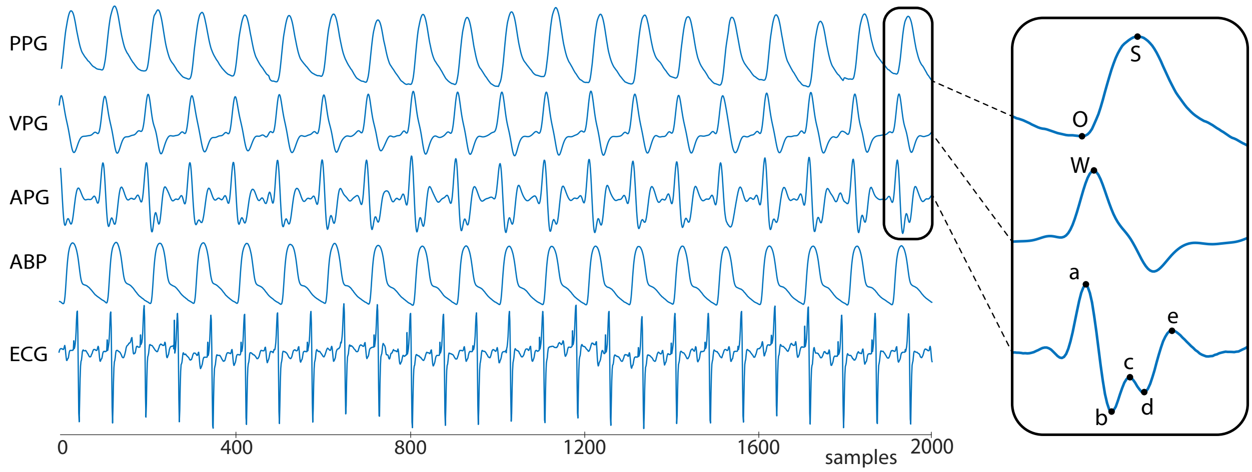

Fig. 1. Fragment of 2000 samples illustrating the morphology of the signals used, together with the representation of the characteristic points of the defined photoplethysmogram (PPG), velocity photoplethysmogram (VPG) and acceleration photoplethysmogram (APG) signals.

II. Material and Methods

Recordings used as dataset were obtained from MIMIC database, a free to use database containing information and biomedical signals recordings from Intensive Care Unit (ICU) patients [7]. The signals used in this study were ECG, PPG and BP (ABP), recorded simultaneously using commercial devices. This records usually present artefacts caused by problems at the time of measurement as loss of contact and movements. The records with irregular morphology, noise or missing data were excluded. Finally, 86 records from different patients distributed in 35 normotensive, 26 prehypertensive and 25 hypertensive were obtained, each one with the three signals used. Signals were recorded simultaneously with a duration of 60 seconds, a common sampling frequency of 125 Hz and a resolution of 8-10 bits [8].

A. Signals Preprocessing

Fourth order Chebyshev II bandpass filter was applied to the raw PPG signal between 0.5 and 10 Hz to improve signal quality and remove minor noises that were not appreciated in previous signal obtaining stage [9]. Furthermore, baseline fluctuation was removed in order to obtain more accurately the PPG feature points amplitudes. Thus, the pulse minimum was setted to zero subtracting the lower envelope from the amplitude of the signal. From this processed PPG signal, photoplethysmographic velocity signal (VPG) and photoplethysmographic acceleration signal (APG) were obtained, being its first and second derivatives respectively [10].

The ABP signals obtained from the database which reflect the change in BP over the cardiac cycle were clear and did not require any processing to be applied. Only the systolic blood pressure (SBP) were detected and used as the BP classification label, being normotensive (<120 mmHg), prehypertensive (120-140 mmHG) and hypertensive (>140 mmHg).

Finally, each ECG segment was processed with a high-pass filter with cutoff frequency of 0.5 Hz to remove the baseline, and a low-pass filter with cutoff frequency of 50 Hz to reduce muscle noise of high frequency and remove power grid interference, being 60 Hz in this case [11]. R-peaks were then detected to obtain the position of each beat along the ECG segment [12].

B. Characteristic Points Extraction

The systolic peaks PPG, VPG and APG 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 as characteristic points and whose amplitude and position in the segment of each signal will be used to obtain the discriminatory features between the different BP levels. The method to obtain this characteristic points in the processed signals has been the search of local maximum and minimum calculated based on thresholds in each of the pulses of the signals. Figure 1 illustrates this characteristic points.

C. Photoplethysmographic and Electrocardiographic features

Recent studies using feature extraction to discriminate between healthy and hypertensive individuals using machine learning classifiers have proposed 13 features. They are three Pulse Arrival Time (PAT) features, peak (R-s), derivate (R-w) and foot (R-O), and 10 PPG features that have obtained the highest correlation with BP levels out of 135 morphological features, including S-c power area divided by O-O power area, b-d and S-c slopes. time spans (S-c, S-d), ratios (c/S, (b-c-d)/a, c/w, d/a) and d amplitude [13]. However, these features may be too specific to the signals used in the study and may not generalise to other individuals.

Therefore, in this study, 23 morphological features selected from various studies that obtain blood pressure level have been proposed and represented in Figure 2. Wave propagation models features as the three PATs and Pulse Transit Time (PTT) are used. On the other hand, PPG, VPG and APG morphology parameters are closely related with BP estimation, thus, S and W amplitude, total area of PPG pulse, areas before and after PPG systolic peak (A1 and A2) and the ratio

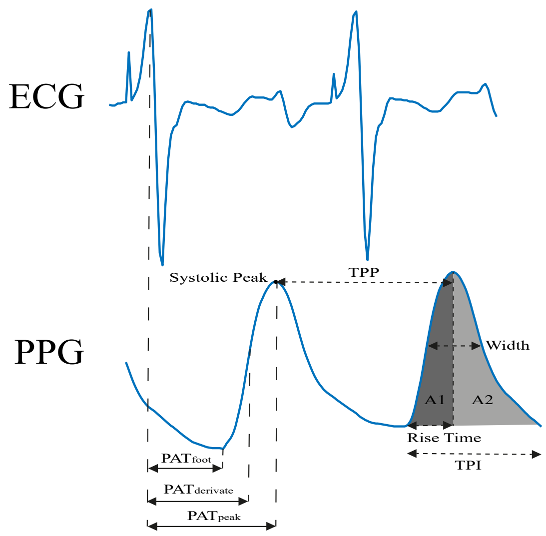

Fig. 2. Representation of main discriminatory parameters used.

between them, intervals between systolic peaks (TPP), total pulse interval (TPI), rise time, pulse width and ratios between the characteristic points of the APG signal(b/a, c/a, d/a, e/a, (b-c-d-e)/a, (b-c-d)/a, (b-e)/a, (c+d-b)/a) were proposed with the aim of improving the discrimination results [10]

After obtaining the values for each discriminatory feature in all data recordings, it is necessary to apply statistical data preparation to manage the missing values and outliers following the absolute deviation from the median (MAD) method [14]. Detected values will be replaced by the median of the values of that parameter in the same hypertension label.

D. Classification models

Classification models as Logistic Regression, AdaBoost Tree, K Nearest Neighbors (KNN) and Bagged Tree used in recent studies were selected because represent the main classification theories [13]. Nevertheless, in this work four alternative classification models have been proposed searching for an improvement in the classification results. Up to 37 different classification strategies such as logistic regression, Naive Bayes, discriminant analysis, support vector machines (SVM), KNN, ensemble classifiers and various types of decision treeshave been tested [15]. Finally, SVM cubic, SVM quadratic, Naive Bayes (Kernel) operator and Coarse Tree were selected as obtained the highest percentages of classificatory accuracy.

E. Statistical analysis

Discriminant features from each record were used as input to train the classification models. Cross-validation leaving one out strategy was carried out applying the classification algorithm once for each record, using all other records as training set and using the selected record as a single test set. Finally, it will be necessary to use statistical tests to evaluate the classification performance. Sensitivity (Se) or ability to detect as positive diseased subjects, specificity (Sp) or ability to detect as negative healthy subjects and F1 score or harmonic mean of accuracy of detecting false positives and sensitivity are used as statistical tests.

III. Results

Since three different groups of patients were labelled, the four most logical paired comparisons between healthy and hypertensive subjects were made, as represented in Table I and Table II. Table I contains the statistical results when predictive parameters that have reported higher correlation with BP levels are used as input for the main classification strategies models. The best classifications performance is obtained comparing normotensive patients with prehypertensive and hypertensive patients in the AdaBoost model with an F1 score of 78%. Other models as Logistic Regression do not discriminate correctly and tend to classify as healthy any subject, as seen in the generalized imbalance between Se and Sp.

Table II presents the improved classification results of this study after applying the proposed photoplethysmographic and electrocardiographic features as input in the alternative classification models. Thus, F1 score value for Naive Bayes and Coarse Tree classification models exceeds 84%.

IV. Discussion

Being able to monitor and detect hypertension with a continuous measurement is of great importance as is the main risk factor for many cardiovascular diseases. Last years, with the growing increase of artificial intelligence techniques, new cuff-less devices have been proposed as an alternative to traditional methods to measure BP. Related studies use PPG signal as its variation in morphology is mainly caused by the activity of the heart and the condition of the vascular walls. Moreover, PPG signals are simple to obtain with nonivasive low cost devices and can be measured in real time.

This studies use PPG and ECG signals, as PAT value is directly related with BP value. This feature used as the only parameter to obtain BP has been studied [6], [16], but the combination with additional PPG characteristics has reported a higher correlation with BP levels [13]. They provide different information, as PAT indicates the transmission of the arterial wave in the blood vessel, whereas PPG morphological features indicate the change in the vascular tissue and blood volume.

Until now, recent studies that have used the PPG signal both to obtain BP level and to classify patients as healthy or hypertensive has not agreed on the predictive features or the classification models to be used, since it depends considerably on the recordings, the database used, patient selection, mode of acquisition, and signal quality. For this reason, this study purpose novel features that have reported improved classification results compared to previous works.

Moreover, the best results have been obtained when normotensive patients are compared with prehypertensive and hypertensive patients. It means that the defined features in prehypertensive patients are more similar to hypertensive patients than normotensive. Furthermore, alerting prehypertensive subjects as diseased is of great interest as generally no

TABLE I

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.

	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 II

PERFORMANCE OF THE FOUR NEW PROPOSED CLASSIFICATION MODELS USING THE NEW CHARACTERISTIC PARAMETERS.

	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%

symptoms are shown in subjects with elevated BP until they are in very advanced stages, so this binary discrimination is the best option for classification.

The main limitations of the study are the low number of recordings, the lack of clinical information related to factors that may imply a higher risk of hypertension such as sex, age or physical condition and sampling frequencies higher than 125 Hz would improve the extraction of characteristic points of the signals.

V. CONCLUSION

This work has proven that the combined analysis of PPG and ECG signals, together with the proposal of new photoplethysmographic and Electrocardiographic Features, as well as the use of alternative classification models, improve the discrimination between healthy and hypertensive individuals. Implement this artificial intelligence methodology in wearable devices may enable the prevention and early diagnosis of hypertension and the associated cardiovascular diseases.

ACKNOWLEDGMENTS

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] T. Unger, C. Borghi, F. Charchar, N. A. Khan, N. R. Poulter, D. Prabhakaran, A. Ramirez, M. Schlaich, G. S. Stergiou, M. Tomaszewski, R. D. Wainford, B. Williams, and A. E. Schutte, "2020 International Society of Hypertension Global Hypertension Practice Guidelines," *Hypertension*, vol. 75, no. 6, pp. 1334–1357, jun 2020.

[2] D. Levy, "The Progression From Hypertension to Congestive Heart Failure," *JAMA: The Journal of the American Medical Association*, vol. 275, no. 20, p. 1557, may 1996.

[3] E. M. Frese, A. Fick, and S. H. Sadowsky, "Blood Pressure Measurement Guidelines for Physical Therapists," *Cardiopulmonary Physical Therapy Journal*, vol. 22, no. 2, pp. 5–12, jun 2011.

[4] K. Kario, "Management of Hypertension in the Digital Era," *Hypertension*, vol. 76, no. 3, pp. 640–650, sep 2020.

[5] S. Kulkarni, "Hypertension management in 2030: a kaleidoscopic view," *Journal of Human Hypertension*, pp. 1–6, nov 2020.

[6] R. Mukkamala, J. O. Hahn, O. T. Inan, L. K. Mestha, C. S. Kim, H. Toreyin, and S. Kyal, "Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 1879–1901, aug 2015.

[7] A. E. Johnson, T. J. Pollard, L. Shen, L. W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, pp. 1–9, may 2016.

[8] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." *Circulation*, vol. 101, no. 23, jun 2000.

[9] Y. Liang, M. Elgendi, Z. Chen, and R. Ward, "Analysis: An optimal filter for short photoplethysmogram signals," *Scientific Data*, vol. 5, no. 1, pp. 1–12, may 2018.

[10] M. Elgendi, "On the Analysis of Fingertip Photoplethysmogram Signals," *Current Cardiology Reviews*, vol. 8, no. 1, pp. 14–25, jun 2012.

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

[12] A. Martínez, R. Alcaraz, and J. J. Rieta, "Application of the phasor transform for automatic delineation of single-lead ECG fiducial points," *Physiological Measurement*, vol. 31, no. 11, pp. 1467–1485, nov 2010.

[13] Y. Liang, Z. Chen, R. Ward, and M. Elgendi, "Hypertension Assessment via ECG and PPG Signals: An Evaluation Using MIMIC Database," *Diagnostics*, vol. 8, no. 3, p. 65, sep 2018.

[14] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764–766, jul 2013.

[15] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge University Press, jan 2013, vol. 9781107057.

[16] Y. Liang, D. Abbott, N. Howard, K. Lim, R. Ward, and M. Elgendi, "How Effective Is Pulse Arrival Time for Evaluating Blood Pressure? Challenges and Recommendations from a Study Using the MIMIC Database," *Journal of Clinical Medicine*, vol. 8, no. 3, p. 337, mar 2019.