The final publication is available at

https://doi.org/10.1109/EHB52898.2021.9657658

Additional Information

Application of Deep Neural Network Models for Blood Pressure Classification based on Photoplethysmograpic Recordings

Jesús Cano[1], Lorenzo Fácila[2], Philip Langley[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, Spain; lfacila@gmail.com
[3] Department of Engineering, University of Hull, United Kingdom; p.langley@hull.ac.uk
[4] Research Group in Electronic, Biomedical and Telecomm. Eng., Univ. of Castilla-La Mancha, Spain; raul.alcaraz@uclm.es

Abstract—The measurement of blood pressure (BP) in an uninterrupted and comfortable way for the subject is essential for early diagnosis and monitoring of cardiovascular diseases (CVD). In fact, hypertension is the main risk factor for CVD because, being a hidden health problem with no symptoms until late stages of the disease are reached. This work investigates whether deep neural network models are able to discriminate between healthy and hypertensive subjects based on photoplethysmographic (PPG) recordings, without the need of electrocardiographic (ECG) recordings as well as avoiding manual morphological feature extraction, as has been popularly used in many previous studies. Recordings analyzed consisted of 635 simultaneous PPG and arterial blood pressure (ABP) signals from 50 different patients. The classification was performed with GoogLeNet, ResNet-18 and ResNet-50 pretrained convolutional neural networks (CNN) using as input images the scalogram of PPG segments obtained by continuous wavelet transformation (CWT). Additionally, Adam and SGDM training solvers were used to compare classification performance. After applying early stopping to avoid overfitting, training was performed with more than half of the epochs using Adam optimizer. ResNet-18 CNN provided the highest classification performance with sensitivity of 95.68 %, specificity of 93.65 %, F1-score of 95.61 % an Area under the Roc area of 98.77 %. Hence, the application of deep neural network classification models using time frequency transformation of PPG recordings has been able to provide outstanding results in blood pressure classification without requiring neither morphological feature extraction nor ECG features.

Keywords—Photoplethysmogram (PPG); Blood Pressure (BP); Deep Learning (DL); Classification Models

I. INTRODUCTION

High blood pressure (BP) is the predominant risk factor for many cardiovascular disease (CVD) as heart failure, atrial fibrillation and stroke among others [1]. Successful BP measurement and control are, consequently, essential for the prevention and early diagnosis of CVD. Conventional cuff-based devices for BP measurement are not designed to provide continuous information throughout the day as need periodic inflation and deflation of the cuff, being uncomfortable and difficult to monitor BP continuously. New wearable blood pressure monitors are expected to enable frequent and accurate measurements with the least stress for the patient in order to detect and manage hypertension in all kind of situations, not only in the office [2].

The most frequently used cuff-less methods to detect and discriminate hypertension combine machine learning techniques with photoplethysmographic (PPG) and electrocardiographyc (ECG) recordings. Features extracted from these signals, as pulse transit time (PTT) and pulse arrival time (PAT), offer acceptable accuracy in Machine Learning based methods [3], [4]. However, synchronized PPG and ECG are required together with accurate feature extraction. Recently, Deep Learning has become a powerful method in image classification problems since convolutional neural networks (CNNs) extracts discriminative and robust features automatically with self learned mechanisms [5]. With the development of neural networks, many CNN such as GoogLeNet [6] and ResNet [7] have emerged. These CNNs are pretrained with ImageNet dataset [8] (more than a million images of 1000 classes) and use transfer learning via feature extraction to reduce the amount of data and computational cost compared to training CNNs from scratch [9].

Recent studies have used image transformation of PPG recordings by continuous wavelet transform (CWT) to generate scalograms and use them as input images to pretrained CNN for BP classification [10], [11]. The aim of this study is to compare the classification performance of GoogLeNet, ResNet-18 and ResNet-50 CNNs to discriminate between the scalogram of PPG waves from normotensive subjects and prehypertensive or hypertensive patients.
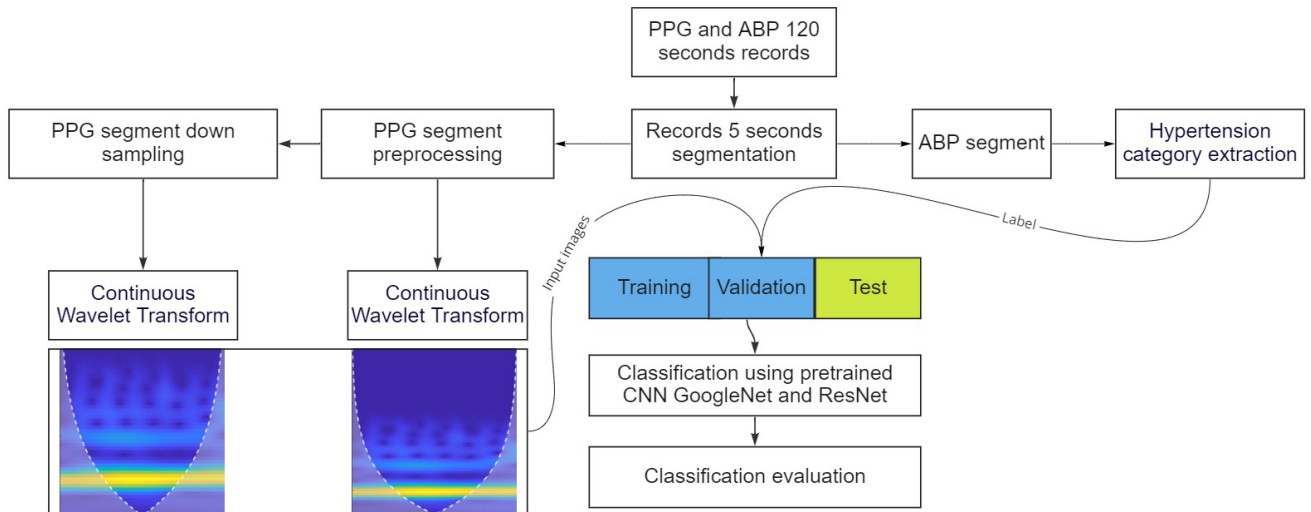
Fig. 1. Block diagram illustrating the deep learning classification methodology employed in the present work.

II. MATERIALS

The dataset was obtained from MIMIC-III Waveform Database [12], containing quasi-continuous recordings of biomedical signals of a single patient throughout an ICU stay. The two signals used were PPG and arterial blood pressure (ABP), which were obtained using commercial devices, so often contain artifacts caused by sensor movements or loss of contact. For this reason, a manual check was carried out to reject recordings with missing peaks, improbable BP values, or signals without their characteristic morphology. Finally, 635 recordings from 50 different patients were selected with simultaneous and stable ABP and PPG signals with 120 s length and 125 Hz sampling rates. In order to introduce this method in the mobile health field, PPG signals are downsampled to 25 Hz, as any morphological feature must be extracted from the signal and its waveform is preserved. Using lower sampling rates reduces the power consumption of the cuff-less device and the amount of data saved and transmitted [13].

III. METHODS

Figure 1 shows an overall block diagram of the whole process applied in this research to PPG recordings, illustrating basic signal processing applied, transformation to images, and CNN classification steps that are going to be described next.

A. Signal Preprocessing

Systolic blood pressure (SBP) was extracted from the mean values of each ABP wave peaks, and were used to target each PPG segments in normotensive (NT) and hypertensive (HT) subjects, with SBP values lower or higher than 120 mmHg, respectively, as defined by the US National Institutes of Health [14]. As most patients have no symptoms in the elevated BP stage, prehypertensive (PHT) subjects (120-140 mmHG) are labelled as HT in binary classification to alert this group as diseased and thus facilitate the prevention and early diagnosis of hypertension. Additionally, PPG signals

were processed with a 0.5-10 Hz Chebyshev II bandpass filter of fourth order to remove noise [15]. After these steps, both signals were cut in 5s length segments, being 15.240 the total number of segments analyzed.

B. PPG Signal Transformation using CWT

Pretrained CNN used in this work accepts RGB images as inputs, thus PPG segments were processed by CWT and transformed to a scalogram, a representation of frequency along the time with the amplitude of the frequency represented by the variation of colours. In this study, it has been used the analytic Morse (3,60) wavelet, setting the VoicesperOctave to 12 to create the CWT. The cone of influence has been included in the scalogram as represents where occur edge effects in the CWT and has obtained better classification results. Finally, RGB images were resized to 224x224x3 to feed the training models, as is a requirement for the classification.

C. Pretrained Convolutional Neural Networks

Three different deep convolutional neural networks were evaluated for the hypertension risk classification problem, GoogleNet [6], ResNet18 and ResNet50 [7] as are the state of art for feature extraction from images and have been pretrained to learn how to extract informative features. In this way, the last learning layer and the final layer of classification are replaced with new layers adjusted to the new training images.

Training options established a minimum Batch size of 128, the validation frequency is modified depending on the number of training images, in this case 76, and the maximum epochs was 25. In order to minimize the effect of overfitting, early stopping technique stops training automatically when the validation loss starts to increase.

D. Classification evaluation

CNN models GoogLeNet, ResNet18 and ResNet50 were evaluated in the classification problem of discriminating between normotensive and hypertensive image representation of

Model	Optimizer	Epoch	Train Loss	Valid Loss	Valid Acc (%)
GoogLeNet	SGDM	1	1.0579	0.8402	57.49
		25	0.2534	0.3594	89.84
	Adam	1	1.6798	2.8843	58.72
		11	0.0850	0.2191	93.37
ResNet-18	SGDM	1	0.8830	0.8403	48.08
		25	0.0331	0.2066	92.96
	Adam	1	0.9322	0.7584	61.06
		10	0.0146	0.2155	94.84
ResNet-50	SGDM	1	0.9429	0.8619	58.76
		22	0.0147	0.2174	92.96
	Adam	1	0.8279	1.0282	54.42
		8	0.0398	0.2131	93.98

downsampled PPG signals. Data images were divided into training and test set using a splitting ratio of 80% and 20%, resulting in 12.192 training images from 508 recordings of 120 segments and 3048 test images from 127 recordings. Training and test recordings were obtained from different patients. In addition, the first group was randomly divided again in 80% for training and 20% for validation, as the validation set was used to prevent models' overfitting.

Additionally, effectiveness of the classification from the models was compared using two different training solvers, Adam [16] and SGDM [17], with the aim to find the best training hyperparameter. Finally, classification performance results of each model with both training solvers were evaluated with different metrics as Accuracy (Acc), F1-score, Sensitivity, Specificity and Area Under the ROC Curve (AUC).

IV. RESULTS

A. Training-Validation Classification

Training performance in terms of training loss, validation loss and validation accuracy obtained by the proposed networks at first and last trained epochs are listed in Table I. Although it was specified in the training options that the maximum number of epochs is 25, it can be seen that in all the models in which the Adam optimizer has been used, the total number of epochs has been between 8 and 11. With this it can be seen that the early stopping technique has performed adequately, thus minimizing overtraining.

Table II shows the classification efficiency when the dataset is randomly divided in training and validation through statistical metrics. It can be seen that all models achieved the highest performance represented with F1-score when the chosen optimizer is Adam. From them, the best classification results are obtained with ResNet-18 with a sensitivity of 95.68%, specificity of 93.65%, F1-score of 95.61% and AUC of 98.77%. Nevertheless, superb results are obtained with all models and optimizers, with F1-score over 91%.

Model	Optimizer	Sen (%)	Spe (%)	F1-score (%)	AUC (%)
GoogLeNet	SGDM	89.38	90.17	91.25	95.88
	Adam	94.42	91.87	94.36	97.51
ResNet-18	SGDM	94.70	90.48	94.04	97.72
	Adam	95.68	93.65	95.61	98.77
ResNet-50	SGDM	94.56	90.67	94.04	97.71
	Adam	94.56	93.15	94.86	98.40

Model	Optimizer	Sen (%)	Spe (%)	F1-score (%)	AUC (%)
GoogLeNet	SGDM	48.97	60.34	54.93	57.04
	Adam	62.84	41.13	60.90	54.98
ResNet-18	SGDM	70.83	51.39	68.51	62.36
	Adam	59.59	55.86	62.00	61.45
ResNet-50	SGDM	69.92	44.06	66.18	57.56
	Adam	59.25	55.86	61.75	59.67

B. Test Classification

Table III shows test results when 20% of the records, representing new patients, are used as input images in the previously trained models. This will test whether the models correctly classify PPG signal representations of patients that were not employed in training and validation sets.

The test results show a significant reduction in the classification performance using new patients, being ResNet-18 with SGDM optimizer the combination that obtained the highest F1- score with 68.51% in addition to sensitivity of 70.83%, specificity of 51.39% and AUC of 62.36%.

V. DISCUSSION

The development of medical devices that provide continuous BP information is of great interest, as current cuff-based devices are not compatible with uninterrupted measurement and monitoring. To this respect, early detection of hypertension would reduce many cardiovascular diseases, as it is its main risk factor.

Liang. Y et al. [4] combined PAT and PPG morphological features to discriminate between NT + PHT vs HT subjects, achieving a F1 score of 88.49%, being lower than our proposed method. Moreover, they were based on manual feature extraction, being difficult to accurately identify feature points, and both PPG and ECG synchronized signals were needed. Our purposed method is simpler, using only PPG recordings that are easily obtained with wearable optical sensors, BP fluctuations are directly reflected in PPG morphology, and the complex manual extraction of Machine Learning features is

not required as discriminant features are automatically and robustly obtained from PPG representation images.

This study purpose was to analyze which of the three most used pretrained CNNs, such as GoogLeNet, ResNet-18 and ResNet-50 with Adam or SGDM training solvers obtain the best classification performance discriminating between normotensive and hypertensive subjects using the CWT of 15.240 PPG segments as input. The classification performance represented in Table II shows outstanding results independently of the neural network and optimizer used, achieving all of them F1-scores higher than 91.25%, although a slight improvement is appreciated using Adam optimizer. On the other hand, training performance comparison shows that the use of Adam optimizer reduces the maximum number of epochs by more than a half, thanks to the implementation of early stopping, that will prevent models from overfitting.

Furthermore, test results are significantly lower than validation results, with no classification model exceeding 70% F1 score. The main reason for this may be due to the fact that cardiovascular dynamics can be a unique feature for each subject. Thus, the relationship between BP and PPG signal and its waveform is not completely generalized for all individuals. Therefore, it has been found that the classification of new individuals that have not been used to train the classification model can be performed with a not so high accuracy, and requires information and signal segments from the same individual in the training, validation and test datasets [18].

Results have shown that the three chosen models with both optimizers are able to classify BP levels with guarantees, being ResNet-18 CNN with the Adam optimizer the one that has achieved the best results by combining the highest values of sensitivity, specificity, F1-score and AUC with a reduction of the number of epochs to 10, avoiding overfitting. Finally, the main contributions of this study compared to related works [10], [19] is the use of subsampled PPG signals which have potential applications in wearable devices as it would reduce the use of memory and computational complexity, the inclusion of PHT subjects to diseased dataset to facilitate early detection of hypertension and the analysis of Adam and SGDM optimizers to asses BP level classification.

VI. CONCLUSIONS

The application of deep neural network classifiers combining continuous wavelet transform of PPG recordings and pretrained CNN models provide superb performance in blood pressure classification. Furthermore, the main features from scalograms are automatically extracted and synchronized ECG recordings are not required for feature extraction. Hence, their employment is encouraged in wearable devices with subsampled PPG signals for blood pressure monitoring as reduce memory and power consumption.

REFERENCES

[1] F. D. Fuchs and P. K. Whelton, "High Blood Pressure and Cardiovascular Disease," *Hypertension*, vol. 75, no. 2, pp. 285–292, feb 2020.

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

[3] X. Ding, B. P. Yan, Y.-T. Zhang, J. Liu, N. Zhao, and H. K. Tsang, "Pulse Transit Time Based Continuous Cuffless Blood Pressure Estimation: A New Extension and A Comprehensive Evaluation," *Scientific Reports*, vol. 7, no. 1, p. 11554, dec 2017.

[4] 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.

[5] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, jan 2019.

[6] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 07-12-June. IEEE, jun 2015, pp. 1–9.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2016-Decem. IEEE, jun 2016, pp. 770–778.

[8] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, dec 2015.

[9] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, jun 2020.

[10] Y. Liang, Z. Chen, R. Ward, and M. Elgendi, "Photoplethysmography and Deep Learning: Enhancing Hypertension Risk Stratification," *Biosensors*, vol. 8, no. 4, oct 2018.

[11] J. Wu, H. Liang, C. Ding, X. Huang, J. Huang, and Q. Peng, "Improving the Accuracy in Classification of Blood Pressure from Photoplethysmography Using Continuous Wavelet Transform and Deep Learning," *International Journal of Hypertension*, vol. 2021, pp. 1–9, aug 2021.

[12] 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.

[13] D. Fujita and A. Suzuki, "Evaluation of the Possible Use of PPG Waveform Features Measured at Low Sampling Rate," *IEEE Access*, vol. 7, pp. 58 361–58 367, 2019.

[14] A. V. Chobanian, G. L. Bakris, H. R. Black, W. C. Cushman, L. A. Green, J. L. Izzo, D. W. Jones, B. J. Materson, S. Oparil, J. T. Wright, and E. J. Roccella, "Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure," *Hypertension*, vol. 42, no. 6, pp. 1206–1252, dec 2003.

[15] Y. Liang, M. Elgendi, Z. Chen, and R. Ward, "An Optimal Filter for Short Photoplethysmogram Signals," *Scientific Data*, vol. 5, no. 1, p. 180076, dec 2018.

[16] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, dec 2014.

[17] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning." PMLR, may 2013, pp. 1139–1147.

[18] J. Esmaelpoor, M. H. Moradi, and A. Kadkhodamohammadi, "A Multistage Deep Neural Network Model for Blood Pressure Estimation using Photoplethysmogram Signals," *Computers in Biology and Medicine*, vol. 120, p. 103719, may 2020.

[19] X. Sun, L. Zhou, S. Chang, and Z. Liu, "Using CNN and HHT to Predict Blood Pressure Level Based on Photoplethysmography and Its Derivatives," *Biosensors*, vol. 11, no. 4, p. 120, apr 2021.