2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Delineator for Arterial Blood Pressure Waveform Analysis Based on a Deep Learning Technique

Nicolas Aguirre[1,2], Edith Grall-Maës[2], Leandro J. Cymberknop[1], Ricardo L. Armentano[1]

***Abstract*— A deep learning technique based on semantic segmentation was implemented into the blood pressure detection points field. Two models were trained and evaluated in terms of a reference detector. The proposed methodology outperforms the reference detector in two of the three classic benchmarks and on signals from a public database that were modified with realistic test maneuvers and artifacts. Both models differentiate regions with valid information and artifacts. So far, no other delineator had shown this capacity.**

I. INTRODUCTION

The analysis of the arterial blood pressure (ABP) constitutes one of the practices to evaluate the state of the circulatory system. This analysis allows the characterization and early detection of cardiovascular pathologies. The ABP morphology (ABPM) is determined by the ventricular ejection pattern and the elastic nature of the arterial tree [1]. Values such as the diastolic and systolic blood pressure (DBP and SBP, respectively) and the occurrence of the dicrotic notch (DN) provide relevant information related to the morphology. ABPM can suffer local alterations, such as those induced by the respiratory rhythm or specific vascular test maneuvers. Additionally, signals are exposed to artifacts that impact the quality of the measurement. The DBP is referenced to the onset of pulse and to the aortic valve opening to blood ejection. The SBP is defined as the occurrence of the maximum pressure value. The DN represents the closure of the aortic valve and is used to calculate the ejection's duration and the start of the diastolic phase. Therefore, the detection of these particular points of ABPM, hereafter called fiducial points (FiP), has to be accurately determined to avoid errors in the evaluation of vascular dynamics.

In previous studies, different delineators have been proposed to be applied to ABPM analysis. Among the most common techniques we can find: weighted slope sum function with adaptive thresholds [2] and ABP feature extraction with logical decision gates [3] for the peak; wavelet-based cascaded adaptive filter [4] for the onsets; analysis of derivatives with logical decision gates [5] for the DN. Only a few studies have presented a transversal analysis [6], [7], [8]. In particular, Pulse Delienator (PUD) proposed by [6] focuses on the analysis of the first ABP derivative, where different logic gates are applied to detect the FiP.

The methodology proposed in this work considers the signals as one-dimensional (1D) images and applies semantic segmentation and data augmentation techniques. Then, simple logical gates are applied to detect the FiP. This methodology analyzes the signal in an offline way.

II. MATERIALS AND METHODS

A. Data Description

The public MIMIC-III database (DB) [9] was used in this study for training the models. This is composed of a myriad of different types of data obtained from patients in intensive-care units. In particular, only ABP records were considered, which were stored at a sampling frequency (S_f) of 125 Hz.

For benchmark purpose, three public DBs with annotations from experts and one other computed from our processing section II-B were used. The first one, called CSL [3] consists of two records of 60 min each one. The records contain annotations of the peak. The second is known as the Fantasia DB (FTS) [10]. It consists of 20 signals of 120 min each one. All signals contain heartbeat annotations from the QRS complex (QRS_c)of the electrocardiogram (ECG). Again, only the records with ABP signals were handled. The third is the Polysomnographic DB (SLP) [11]. It consists of approximately 80 h of ABP and ECG recordings of 16 subjects with annotations of the beats in the ECG channel. As the architectures presented below were trained for 125 Hz, all signals were resampled to that frequency. Finally, 15% of the dataset available after the processing stage was used as the fourth DB. This subset will be called MIMIC processed dataset (MPD), as can be observed in Fig. 1. Data augmentation techniques, described below, were only applied once to the MPD.

As a summary, CSL was used to evaluate the peaks detection directly. FTS and SLP were used for the onset,

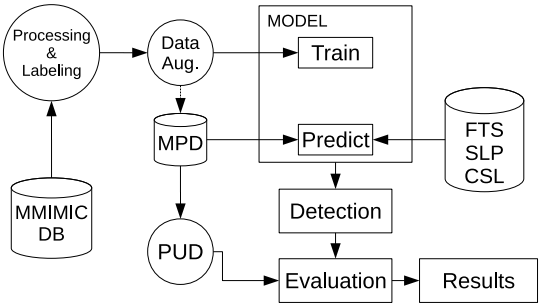

Fig. 1. Proposed methodology

This work was partially supported by the Universidad Tecnologica Nacional (UTN) (Grant: ICUTIBA7647) and by the ML-Cardyn project, co-funded by the European Union. The authors would like to thank Europe for its commitment in Champagne-Ardenne with the FEDER.

[1]Aguirre, N., Cymberknop, L. and Armentano, R. are with GIBIO, UTN, Buenos Aires, Argentina nicolas.aguirre@utt.fr

[2]Edith Grall-Maës and Aguirre, N. are with LIST3N, Université de Technologie de Troyes, Troyes, France edith.grall@utt.fr

and evaluation was done indirectly through ECG annotations. Finally, MPD was used to evaluate the whole of the FiP.

B. Processing and Labeling Description

Only records containing ABP signals longer than 10 min were analyzed from the MIMIC DB to ensure signal stability. For each signal, two 15-s segments separated by a 5-min interval were analyzed. For each beat, the temporal occurrences of the FiP were recorded. Due to the lack of fully labeled ABPMs, PUD [6] was used to detect the FiP and to define the labels of each signal point.

The following criteria were used to ensure good quality in the segments. First, segments constituted by 5% of constant values (no signal info) or peaks (saturation in sensor) were discarded, as in [12]. Then, the characteristics of each heartbeat were analyzed. The SBP was limited between 80 and 200 mmHg and the DBP to 50 and 120 mmHg. The duration of one heartbeat was limited from 62.5 to 187.5 ms. The skewness index was set to be greater than 0 to avoid unrealistic ABPM [13]. A heartbeat was marked when it did not meet any of these requirements. If more than half of the beats in a segment were marked, the segment was discarded. Finally, if a segment had more than 7 mmHg difference between the average DBP and the minimum DBP, the whole segment was also discarded. The same restriction was applied to SBP, but with a difference of 30 mmHg. When a segment was discarded, the next two segments were analyzed. If no pairs of good quality segments were found in the record, it was discarded and the next one was evaluated. In the next step the signal points were classified depending on the heartbeat interval to which they belong. Thus, the classes [*onset* - *peak*] (OP), [*peak* - *DN*] (PD) and [*DN* - *onset*] (DO) were defined. The regions with beats that were marked in the previous step but were not enough to discard the segment were classified as [*artifact*] (AR).

C. Data Augmentation Description

Data augmentation is a technique that consists of applying specific random transformations on the data before using it. The implemented variations were of two types: the replacement of beats by artifacts and changes in the ABP baseline.

In the first type, up to 20% of the beats were randomly removed and the class of points was changed to AR. Then, values of the signal corresponding to the regions of AR class were replaced by a constant value K_1 or by a straight line that joins the previous beat to the one eliminated with the next one. K_1 values were taken from a random uniform distribution $\mathscr{U}$(-0.5, 0.5). Finally, to the replaced points it was added a Gaussian noise $\mathscr{N}(0, \sigma_{artifact})$, where $\sigma_{artifact}$ was randomly sampled from a $\mathscr{U}$(0.01, 0.5)

Independently to the replacement of beats by artifacts, the second type of variation was added. This consisted of adding a sigmoid function $\sigma(x)$, a sinus function $\sin(x)$, or making no change at all. These three cases had the same probabilities. $\sigma(x)$ was sampled in the range [-5, 5] and was used to represent an increase or decrease in ABP baseline. It can be observed in specific test, as in the

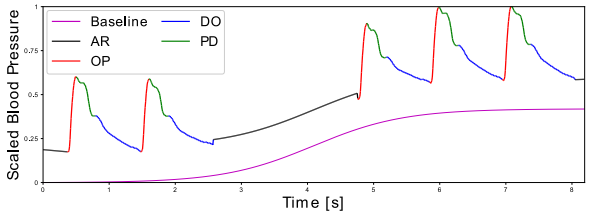

Fig. 2. Signal example applying data augmentation techniques. The OP, PD, DO and AR classes are colored in red, green, blue, and black respectively. The scaled Valsalva's maneuver $\sigma(x)$ is shown in magenta.

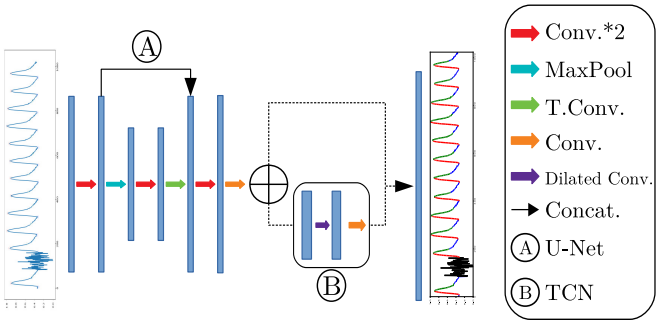

Fig. 3. Neural network architectures. The dashed arrows mean for the two possible models. Each model ends with a point-wise softmax function.

Valsalva's maneuver [14], where the baroreflex sensitivity, an indicator of various pathologies, is revealed [15]. To cover a larger number of scenarios, $\sigma(x)$ was scaled by a constant value K_2 randomly sampled from $\mathscr{U}$(0.2, 0.5) and multiplied by 1 or -1, increase or decrease, respectively. $\sin(x)$ was used to represent the variations in ABPM produced by the respiration. Phase, frequency and amplitude were randomly sampled from $\mathscr{U}$(0, 2π), $\mathscr{U}$(0.1, 0.3) and $\mathscr{U}$(0.05, 0.1) respectively. Finally, signals were scaled in the range [0, 1].

Data augmentation techniques were randomly repeated during the training while, according to Fig. 1, they were applied once to create the MPD. Only 8.192 s (1024 time-steps) after the first second were taken from each signal to generate MPD, and the remaining were discarded. An example of data augmentation techniques can be seen in Fig. 2. The noise was not added for visual appreciation.

D. Neural Network Architecture

Semantic segmentation is a technique widely applied to images. This approach was implemented to 1D ABP signals. Two models were trained to classify each signal's point based on the work of [16]. The first one is a traditional U-Net. The second is the U-Net with an 8-layer temporal convolutional network [17] (TCN), hereafter called U-TCN. The architectures are summarized in Fig. 3.

The U-Net involves an encoder-decoder with skip connections. At the end of each encoder level, the signal is compressed with a maxpool operation and the decoder expands it again with a transpose convolution (T.Conv) operation. The decoder generates specific features to each level from the feature map of the previous level and the corresponding connection with the encoder. In our architecture, each level of the encoder and decoder contained two blocks of convolution, batch normalization and ReLU activation (Conv.*2).

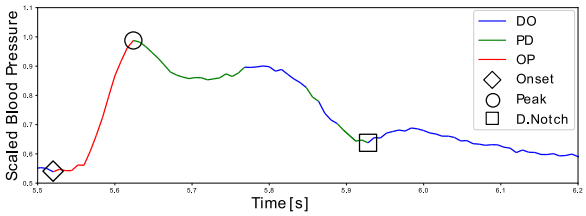

Fig. 4. Detection method. The black shapes are the fiducial points detected. Blue, green, and red lines represent the predicted classes *dicrotic notch - onset*, *peak - dicrotic notch*, and *onset - peak*, respectively.

TCN's architecture was developed to retain sequence information of any length. The retention is achieved with two techniques. The first one is the use of dilated convolutions that exponentially increase the reception field of each layer. The second is the use of residual blocks. This effectively allows layers to learn modifications to the identity mapping rather than the entire transformation, which has been shown to benefit deep neural networks [17].

Outputs from the models were obtained from a point-wise softmax function over the final feature map of size equal to four (one per class). As a penalization, the cross-entropy loss function was used. To update the models the Adam optimizer was used. The learning rate value was set to 0.001. Each model was trained for 20 epochs with the 72,5% from the processed MIMIC DB, leaving 12,5% to validation. The remaining 15% is the aforementioned MPD dataset.

E. Point Detection

Following the point classification, a method was implemented to mark the FiP. First, potential FiP are marked when there is a meaningful change of class in the classification. For example, only the points that were classified as DO and that were preceded by a PD point were marked as a potential DN event. The only exception was the onset event, which could also be preceded by a point classified as AR. Second, to solve the multiple marks of DN along a single beat and improve the results, a backward scan of the signal was done. Only the previous marked points that follow the logic temporal sequence: "*end of pulse - dicrotic notch - peak - onset*" were considered. The "*end of pulse*" could be either a point classified as OP or AR. Finally, for each heartbeat only the first DN marked was kept in a backward sense. Black shapes in Fig. 4 illustrate the detection method.

III. EVALUATION AND RESULTS

For evaluation purpose, two classic benchmark parameters were observed: sensitivity (Se) and positive predictivity (P^+):

$$Se = TP/(TP+FN); \qquad P^+ = TP/(TP+FP)$$

where TP, FN and FP refer to number of true positive, false negative, and false positive, respectively. Additionally, an error rate $E^{\%}$ was also evaluated.

$$E^{\%} = (FP+FN)/(TP+FP)$$

As the models had a fixed input size, the predictions were done by batches and then were rejoined. The detection's evaluation with respect to CSL, FTS and SLP is summarized in Table I. As mentioned, in the FTS and SLP only are available QRS_c annotations. To simplify the analysis, it will be omitted the occasional discrepancy between QRS_c annotations and effective ABPM [7]. A TP was considered if between two QRS_c annotations were one detection. But if between two consecutive QRS_c more than one was detected, the first was considered TP and the others FP. If none was found, it was counted as a FN. The lower $E^{\%}$ for FTS was 1.30% with the U-TCN model, while for SLP was 0.83% with PUD. Particularly, as the CSL data allow direct detection on the peak, a 8-ms windows was determined. In this case, the lower $E^{\%}$ was 0.19% with the U-TCN model.

A smaller window of 3-ms was taken for a transversal evaluation of all FiPs with the MPD set. If multiple points were marked in the 3-ms window, one was considered TP and the others as FP. Results are given in Table II. The lower $E^{\%}$ for the onset, peak, and DN were 0.69%, 0.49%, and 1.98% respectively, reported by the U-Net model. It was followed by the U-TCN where the $E^{\%}$ for the onset, peak, and DN were 0.80%, 0.53%, and 2.01% respectively. The $E^{\%}$ with PUD for the onset, peak, and DN were 19.54%, 19.25%, and 21.06% respectively. The FP quantities due to regions of artifacts for each FiP in PUD were 6033. Hence, the FP on PUD due to the built-in baseline variations were 2996, 2885, and 3668 for the onset, peak, and DN, respectively.

TABLE I

PERFORMANCE ON COMMON BENCHMARK DATABASES

Data	Model	TP	FP	FN	Se	P^+	$E^{\%}$
FTS	PUD	135748	24	2082	98.29	99.98	1.56
	U-Net	142806	343	1573	98.91	99.76	1.34
	U-TCN	142833	320	1546	98.93	99.78	**1.30**
SLP	PUD	315823	17	2589	99.19	99.99	0.83
	U-Net	365695	1368	2651	99.28	99.63	1.09
	U-TCN	365996	1377	2350	99.36	99.63	1.01
CSL	PUD	13055	21	24	99.82	99.84	0.34
	U-Net	13052	0	27	99.79	100.00	0.21
	U-TCN	13054	0	25	99.85	100.00	**0.19**

Se, P^+ and $E^{\%}$ in percentage.

TABLE II

EVALUATION ON MIMIC PROCESSED DATASET

Point	Model	TP	FP	FN	Se	P^+	$E^{\%}$
Onset: 64133	PUD	59727	9029	4406	93.13	86.87	19.54
	U-Net	63810	118	323	99.50	99.82	**0.69**
	U-TCN	63798	179	335	99.48	99.72	0.80
Peak: 64216	PUD	59888	8918	4328	93.26	87.04	19.25
	U-Net	64042	140	174	99.73	99.78	**0.49**
	U-TCN	64017	138	199	99.69	99.78	0.53
D.N. 64155	PUD	59012	9701	5143	91.98	85.88	21.60
	U-Net	63465	580	690	98.92	99.09	**1.98**
	U-TCN	63485	618	670	98.96	99.04	2.01

D.N. is to dicrotic notch. Se, P^+ and $E^{\%}$ in percentage.

IV. DISCUSSIONS AND FUTURE WORKS

Despite the wide variety adopted by ABPMs, the presented methodology was successfully applied, where both FiP and artifacts were adequately differentiated. Additionally, this methodology has the potential to further improve its per-

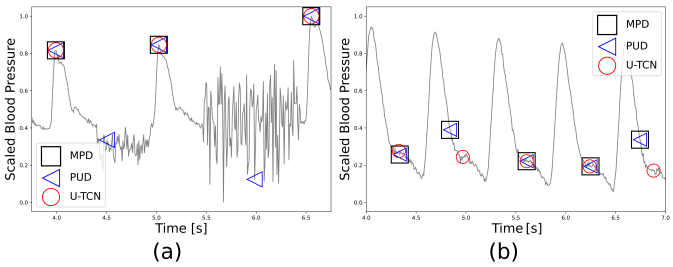

Fig. 5. Comparison between Pulse Delineator (PUD) and U-TCN model for MIMIC Processed Dataset (MPD). (a) Peak detections. (b) Over-penalization examples for the detection of dicrotic notches.

formance in terms of the incorporation of a wide variety of ABPM during training phase. This is a limitation that seems to have the algorithms based on signal processing. For them, the use of logical gates and thresholds seems to be the way to increase the range of ABPM where they success. However, the proposed methodology requires labeled data.

From Table I it can be observed that at least one of the proposed models leads the $E^{\%}$ performances in terms of FTS and CSL. Despite that for SLP the lower $E^{\%}$ was achieved by PUD both models showed similar performances. It is important to note the differences in total annotations (FP + FN) between PUD and our results on FTS and SLP. This is because for this work, a visual check of all the signals and annotations was not performed as asserted by [6]. Nevertheless, it is pertinent to highlight that in general, all the models had fewer FN quantities than PUD. Therefore, the presented methodology results in greater sensitivity to the onset. In summary, although based on the response of the work by [6] to get the point labels, our approach proved to be, in accuracy terms, competitive with the state of the art in three classical benchmark DBs.

The MPD set was used for a transversal evaluation of all the FiP. Results in Table II show a better performance than PUD. Even without accounting the errors caused on artifact regions, PUD showed to be less regular in the detection to realistic modifications made in section II-C. It can be seen a large improvement in the FiP detection for both models, compared with the reference delineator PUD. Some results with model U-TCN can be seen in Fig. 5. Furthermore, for the models proposed, it was observed that in some signals the FN and FP accounts were over-penalized, in particular for DN detection. That happened when the labeling in section II-B was found to be incorrect. The generalization property from statistical learning encourages correct classifications, even though they were counted as incorrect later. A case of this is shown in Fig. 5(b). In summary, even ignoring the error due to artifact regions and the over-penalization on DN, the methodology presented was superior, despite the models.

Results show that batch analysis does not affect detection performance. Furthermore, having enough memory to train the models, the same methodology could be applied for a higher S_f and longer input size. On the other hand, simple custom penalty functions [16] were explored and three additional classes were added, each one specific to the FiP. This allowed us to remove the step detailed in section II-E. Nevertheless, due to the class imbalance, this method could not be fully implemented although it could be considered as future research.

V. CONCLUSIONS

In this paper a new methodology for detecting FiP in the ABPM is presented. Unlike other works where different signal processing techniques and intricate logic gates are used, our work is based on deep learning techniques and naive logic gates. Our methodology also differentiate regions of common artifacts. It could be observed that the detection of FiP is accurate even in the context of simulated maneuvers such as Valsalva. A dataset randomly affected with artifacts and realistic modification of the ABP baseline, but without affecting the temporal occurrences of the FiP, was generated. In this scenario, our approach proved to be efficient in terms of errors during detection.

REFERENCES

[1] A. P. Avolio, *et al.*, "Arterial blood pressure measurement and pulse wave analysis-–their role in enhancing cardiovascular assessment," *Physiological Measurement*, vol. 31, no. 1, pp. R1–R47, Nov. 2009.

[2] W. Zong, *et al.*, "An open-source algorithm to detect onset of arterial blood pressure pulses," in *Computers in Cardiology, 2003*, Sept. 2003, pp. 259–262.

[3] M. Aboy, *et al.*, "An automatic beat detection algorithm for pressure signals," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 10, pp. 1662–1670, Oct. 2005.

[4] L. Xu, *et al.*, "Baseline wander correction in pulse waveforms using wavelet-based cascaded adaptive filter," *Computers in Biology and Medicine*, vol. 37, no. 5, pp. 716–731, May 2007.

[5] M. J. Oppenheim and D. F. Sittig, "An Innovative Dicrotic Notch Detection Algorithm Which Combines Rule-Based Logic with Digital Signal Processing Techniques," *Computers and Biomedical Research*, vol. 28, no. 2, pp. 154–170, Apr. 1995.

[6] B. N. Li, *et al.*, "On an automatic delineator for arterial blood pressure waveforms," *Biomedical Signal Processing and Control*, vol. 5, no. 1, pp. 76–81, Jan. 2010.

[7] A. Pachauri and M. Bhuyan, "Wavelet Transform Based Arterial Blood Pressure Waveform Delineator," *INTERNATIONAL JOURNAL OF BIOLOGY AND BIOMEDICAL ENGINEERING*, Feb. 2012.

[8] O. Singh and R. K. Sunkaria, "Detection of Onset, Systolic Peak and Dicrotic Notch in Arterial Blood Pressure Pulses," *Measurement and Control*, vol. 50, no. 7-8, pp. 170–176, Sept. 2017.

[9] A. E. W. Johnson, *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, p. 160035, May 2016.

[10] N. Iyengar, *et al.*, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *Am J Physiol Regul Integr Comp Physiol*, vol. 271, no. 4, pp. R1078–R1084, Oct. 1996.

[11] Y. Ichimaru and G. B. Moody, "Development of the polysomnographic database on CD-ROM," *Psychiatry and Clinical Neurosciences*, vol. 53, no. 2, pp. 175–177, 1999.

[12] G. Slapničar, *et al.*, "Blood Pressure Estimation from Photoplethysmogram Using a Spectro-Temporal Deep Neural Network," *Sensors*, vol. 19, no. 15, p. 3420, Jan. 2019.

[13] M. Elgendi, "Optimal Signal Quality Index for Photoplethysmogram Signals," *Bioengineering*, vol. 3, no. 4, Sept. 2016.

[14] D. L. Eckberg, "Parasympathetic cardiovascular control in human disease: A critical review of methods and results," *Am J Physiol Heart Circ Physiol*, vol. 239, no. 5, pp. H581–H593, Nov. 1980.

[15] K. J. Osterziel, *et al.*, "Baroreflex sensitivity and cardiovascular mortality in patients with mild to moderate heart failure." *British Heart Journal*, vol. 73, no. 6, pp. 517–522, June 1995.

[16] O. Ronneberger, *et al.*, "U-net: Convolutional networks for biomedical image segmentation." [Online]. Available: http://arxiv.org/abs/1505.04597

[17] S. Bai, *et al.*, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," *arXiv:1803.01271 [cs]*, Apr. 2018.