Evaluation of Cuff-less Blood Pressure Monitoring Models over Multiple Data Sets

Jim Barnes, Colum Crowe, Brendan O'Flynn, Salvatore Tedesco*

***Abstract*—Blood pressure (BP) monitoring via cuffless devices has gained significant attention in the last few years. Despite a plethora of works having been produced in this field based on traditional machine learning (ML) or deep learning (DL) models, very limited research has been carried out in terms of the external validation and reproducibility of said models to ensure that they are of clinical use. To the best of the authors' knowledge, this is the first study to evaluate several of the currently most well cited ML/DL-based models for cuffless BP monitoring over multiple independent data sets. The results of this investigation in reproducibility are reported with particular recommendations provided regarding standardized data collection protocols, models and signals, data recording length, and open access data as potential steps to overcoming the challenge of reproducibility in ML/DL models in this field and the health domain in general.**

Keywords—cuff-less blood pressure, blood pressure monitoring, reproducibility

I. Introduction

High blood pressure (BP), known as hypertension, is a strong predictor of cardiovascular diseases (CVD) [1]. There are many different methods of measuring BP, with varying levels of accuracy. The current gold-standard is intra-arterial BP monitoring [2]. This procedure works by inserting a cannula needle into a suitable artery. The cannula is also connected to a fluid filled system, which is connected to an electronic patient monitor. This means that BP is able to be monitored constantly, with each heartbeat. It is most often used when a patient is in intensive care as it is an invasive procedure. However, in most cases such a high level of accuracy is not needed, with the British Hypertension Society (BHS) protocol giving the highest possible A grade to BP measurement devices with 60% of their predictions within an error of 5mmHg, 85% of predictions within 10mmHg and 95% of predictions within 15mmHg [3].

The auscultation method is a non-invasive method of measuring BP [4]. This technique was previously widely used, but it has been replaced by the oscillometric method in many cases. The oscillometric method [4], while also using a cuff, does not necessitate a quiet environment and is less reliant on the skill of the clinician as more training is needed to hear the different Korotkoff sounds.

This work was supported in part by the Science Foundation Ireland which are Co-Funded through the European Regional Development Fund under Grant 12/RC/2289-P2-INSIGHT and Grant 16/RC/3918-CONFIRM. Aspects of this publication have also emanated from research supported by a research grant from the Disruptive Technologies Innovation Fund (DTIF) project HOLISTICS funded by Enterprise Ireland (EI).

All authors are with the Tyndall National Institute, University College Cork, Lee Maltings Complex, Dyke Parade, T12R5CP Cork, Ireland.

* Corresponding author. e-mail: salvatore.tedesco@tyndall.ie

However, both techniques offer only a once-off measurement rather than continuous, and continuous BP measurements would allow the investigation of transient changes in BP which may give insights into mechanisms of BP control and patients' management [5]. Recent advances in non-invasive estimation of BP have been focused on cuffless approaches generally relying on the electrocardiogram (ECG) signal and photoplethysmogram (PPG) waveform [6]. While there is a considerable amount of previous literature [7-8] in the last decade which has investigated the feasibility of real-time cuffless continuous BP estimation, very little research has been carried out on validating those models outside of the original data sets that were used for building them. It is essential that these algorithms, which generally rely on machine learning (ML) models based on features extracted from ECG and/or PPG signals or deep learning (DL) models using the raw time series, would perform equally well on multiple independent data sets for clinical usefulness to be guaranteed. Despite the numerous reviews published in the area of cuffless BP monitoring [7-8] and the establishment of standardized clinical validation protocols for cuffless BP devices [9], very little research has been done on the validation of the solutions currently available. For instance, in a systematic review by Islam et al. [10] it was shown that the accuracy of wearable cuffless BP devices varied substantially based on different sensing technologies and validation approaches adopted, thus limiting their direct comparison. Moreover, the majority of the studies analyzed in [10] did not provide detailed descriptions of the algorithms for BP measurement.

To the best of the authors' knowledge, this is the first study that evaluates several ML/DL-based models for cuffless BP monitoring over multiple independent data sets. The objective of this study is therefore to evaluate if those supervised models have the actual potential to be used in current clinical settings. For this analysis, we will be focusing on methods that use either a combination of an ECG and PPG, or PPG exclusively, as it has been shown that PPG alone can be used to predict hypertension [11], even though the majority of studies rely on both ECG and PPG [12]. The manuscript is structured as follows: Sections II and III presents the data sets and models investigated in the study; Section IV illustrates the results of each approach with Section V discussing those results and their impact and Section VI drawing the conclusions.

II. Data sets

Table I summarizes the available data sets adopted in this investigation showing details regarding the number of

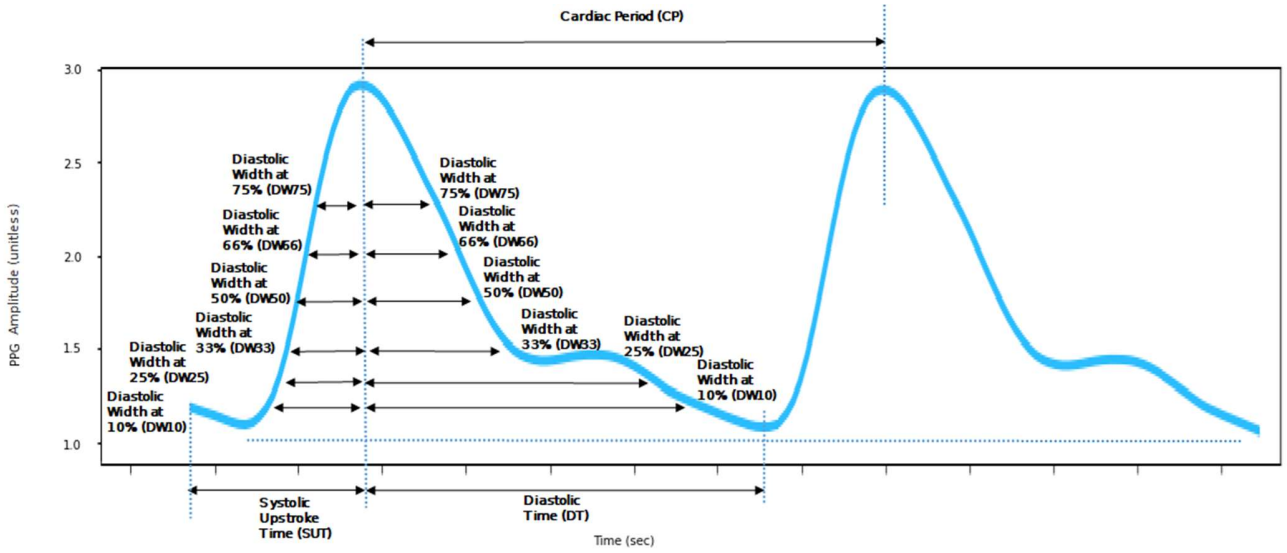

Figure 1. Features description as indicated in [18]

subjects included, presence of ECG and PPG sensing signals, and information about the BP measurements to be used as ground truth provided either in the shape of raw time-series signals (for both systolic and diastolic BP) or as discrete once-off measurements. These details had an influence on which models could be used on each data set. For example, any model (such as the ones based on DL) requiring a full BP waveform as a target label could not be used with data sets which only have once-off measurements, nor could models requiring ECG signals be applied to data sets with only PPG data. For this reason, not all data sets could be used on every model being recreated. Given the large diffusion of the MIMIC database in this field, it was decided not to consider MIMIC in this investigation so as to have a better understanding of how those models behave on datasets as varied as possible. Moreover, also the related Perfusion Signal Quality Index (PSQI) is included in Table I.

TABLE I. DATA SETS CHARACTERISTICS

Data set	Data set full size	PPG	ECG	BP measurements	PPG Perfusion Signal Quality Index
Kachuee et al. [13]	942 subjects	Yes	Yes	Yes, signal	237
UQVSD [14]	32 subjects	Yes	Yes	Yes, systolic and diastolic signal	310
Liang et al. [15]	219 subjects	Yes	No	Yes, discrete	246
Aguirre et al. [16]	5854 subjects	Yes	No	Yes, signal	384
Mehrgardt et al. [17]	22 subjects	Yes	Yes	Yes, discrete	1175

III. MODELS EMPLOYED

Table II summarizes the models considered in this investigation. These models have been chosen as they have shown promising results in the original studies and there has been sufficient information provided for their reproduction. For both Lo et al. [19] and Ibtehaz et al. [20], the code was made available online by the authors, whereas for Kurylyak et al. [18], Li et al. [21], and Athaya et al. [22] feature extraction and models were recreated based on the descriptions provided.

TABLE II. MODELS CHARACTERISTICS

Authors	Model	Code
Kurylyak et al. [18]	Feed-forward neural network, with 35 neurons and 20 neurons in the two hidden layers plus input and output layers	Manual reproduction
Lo et al. [19]	Long-Short Term Memory neural network with 128 neurons in each hidden layer	Made available by authors
Ibtehaz et al. [20]	Approximation network made of a U-Net model consisting only of convolutional layers, followed by a refinement network (MultiResUNet), all consisting of convolutional layers	Made available by authors
Li et al. [21]	One bi-directional LSTM layer followed by four LSTM layers with residual connections between layers	Manual reproduction
Athaya et al. [22]	5-layer U-Net model	Manual reproduction

Each data set was split into a training and testing set with a 80/20 ratio in each model reproduction. A model is re-trained on each new data set before being tested. The training and testing sets in each paper are completely disjointed. Where possible, the models were re-evaluated also on the dataset used to train the model in the original study. All the models were run using an Intel(R) Core(TM) i7-4600U CPU @ 2.10GHz 2.70 GHz processor.

Detailed information about each reproduced model are illustrated below.

Kurylyak et al. [18]

This study deals with BP prediction using exclusively PPG signals, as it avoids possible synchronization issues between ECG and PPG as in some cases in sections of the MIMIC database [23]. This paper uses 21 features, mostly

relying on diastolic width (DW) and systolic width (SW), from each PPG cycle to predict BP. A graphical description of those features is shown in Figure 1. Along with these features, synchronized systolic BP (SBP) and diastolic BP (DBP) are extracted for each PPG cycle and used as the target label for training. The model consists of a feed-forward dense neural network with two hidden layers (with 35 and 20 neurons, respectively) and an output layer with two neurons to predict both the SBP and DBP simultaneously. The model was built relying on the MIMIC database.

Lo et al. [19]

This paper proposes using both ECG and PPG to estimate BP. The inclusion of the ECG helps with the measurement of PTT which is correlated with BP [24]. Both these signals are combined into a two-dimensional sequence and only the subsequence containing the latest k points is used to estimate the latest point of the blood pressure. The signals are then fed into a Long Short-Term Memory (LSTM) neural network consisting of two hidden layers. The data set by Kachuee et al. [13] was adopted for the implementation of the model by the authors.

Ibtehaz et al. [20]

This study is based on PPG signals exclusively. As with Lo et al.'s paper, the entire raw PPG signal is inputted into the model to predict SBP and DBP. Each subject's data was segmented into signals of 8 seconds. After pre-processing, the signals are fed into an approximation network, e.g., a U-Net model with deep supervision using only convolutional layers. The output of the network is then fed into a refinement network (called MultiResUNet), also based on a U-Net architecture, which includes Multi-Residual blocks and Residual paths, and has the goal to fine tune the results of the approximation network into a more accurate prediction. The authors used a subset of the MIMIC-III data set compiled by Kachuee et al. [13] for model development.

Li et al. [21]

The methodology set forth in this paper is based on the extraction of seven features from both ECG and PPG signals to be used as input into the proposed ML model. The extracted features are:

- PTT, which is the distance between the peak of the ECG and max slope of the PPG;
- Heart rate, the time between consecutive ECG peaks;
- Reflection index, which is b/a, where a is the amplitude of the systolic peak and b is the difference in amplitude between systolic peak and the second peak following the dicrotic notch;
- Systolic timespan, which is tn_n-tf_n, the time difference between the dicrotic notch and the foot of the cycle;
- Up time, which is tp_n-tf_n, e.g., the time difference between the systolic peak and the foot of the cycle;
- Systolic volume, which is the integral of the PPG signal from tf_n to tn_n, e.g., between the foot of the current cycle and the dicrotic notch;
- Diastolic volume, which is the integral of the PPG signal from tn_n to tf_{n+1}, e.g., between the dicrotic notch and the foot of the following cycle.

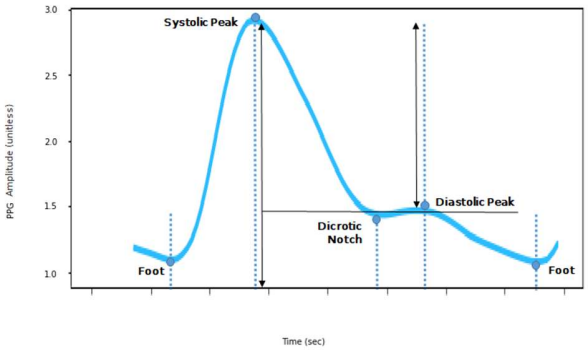

Figure 2. PPG signal as shown in [21]

A graphical description of those features is shown in Figure 2. A cleaned and preprocessed subset of the MIMIC-II data set as prepared by Kachuee et al. [13] was used in the original paper. The authors developed a deep learning architecture consisting of a bidirectional layer of LSTM as the first layer followed by four layers of LSTM with residual connections added between each of those layers.

Athaya et al. [22]

The model developed by Athaya et al. [22] is based on a modified version of the U-Net architecture (Figure 3). This model adopts the PPG signal only and requires a BP waveform to be used as the ground truth. Data is initially filtered using an equiripple FIR filter and normalized, then segmented into windows with a length of 350 samples and 100 sample overlap. Signals were then phase-shifted and finally trimmed into windows of 256 samples, before being inputted into the model. The model was built based on two combined data sets, MIMIC and MIMIC-III.

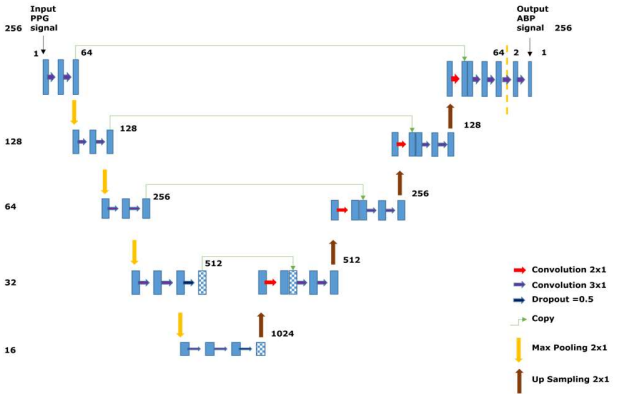

Figure 3. Model architecture as depicted in [22]

IV. RESULTS

TABLE III. RESULTS – MEAN ABSOLUTE ERROR (MMHG)

	Original data set	MAE SBP - DBP	Tested data sets	Data set size	MAE SBP	MAE DBP
Kurylyak et al. [18]	MIMIC - 15,000 beats	3.8 - 2.2 ([18])	Kachuee et al. [13]	159 subjects, 15,284 beats	7.7	2.6
			UQVSD [14]	29*540s recordings, 15,165 beats	17.1	6.1
			Liang et al. [15]	219 subjects, 3×2.1s recordings, 1,242 beats	17.2	9.1
			Aguirre et al. [16]	793 subjects, 15s recordings, 15,010 beats	17.8	7.2
			Mehrgardt et al. [17]	8 subjects, 3×360s recordings 15,067 beats	9.2	5.1
Lo et al. [19]	Kachuee et al. [13] - 84 subjects	2.7 - 1.6	Kachuee et al. [13]	84 subjects	6	3.4
Ibtehaz et al. [20]	Kachuee et al. [13] – 942 subjects	5.7 - 3.5	Aguirre et al. [16]	580 subjects	18.2	12.1
			Aguirre et al. [16]	5,800 subjects	23.4	9.4
Li et al. [21]	Kachuee et al. [13] - 50 recordings, 6,852 beats / Kachuee et al. [13] – 3,000 recordings, 678,202 beats	0.73 - 0.55 / 6.7 - 2.5	UQVSD [14]	25 subjects, 19,875 beats	12.4	11.8
			Mehrgardt et al. [17]	22 subjects, 44,570 beats	9.8	6.7
Athaya et al. [22]	MIMIC - 100 subjects	3.68 - 1.97 ([22])	Kachuee et al. [13]	100 subjects	6.8	3.5
			Aguirre et al. [16]	100 subjects	7.6	4.9

Models described in Section III have been reproduced and results are summarized in Table III in terms of Mean Absolute Error (MAE) and discussed separately below.

Kurylyak et al. [18]

The model was recreated using five different data sets to compare with the original. As evident from the results, the features extracted and the adopted architecture are not as effective in predicting BP when used on a different data set. A possible explanation for this is that the feature extraction process in [18] relies on ideal PPG waveforms (e.g., Figure 1). PPG signals collected in other data sets are not as clean as the ones used in [18] as measurements can be noisy and present high variability between individuals making it difficult to discern the different sections of the cycle. Figure 4 shows an example of PPG signals from other data sets used in this analysis with more noise present. Therefore, it is reasonable to believe that the decrement in accuracy is due to measurements in which the signal largely deviates from the ideal waveform.

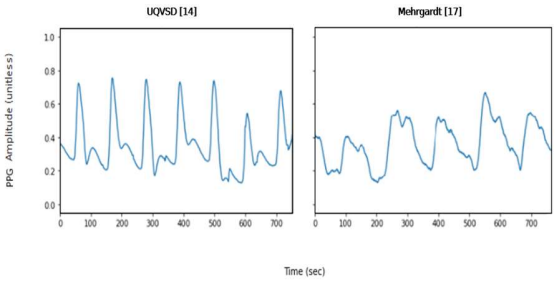

Figure 4. Example of noisy PPG signals from [14] and [17]. Signals were filtered and normalised. It is evident how PPG signals could vary across different datasets

Lo et al. [19]

The model was built from the Kachuee et al. [13] data set and specifically requires ECG and PPG time-series as input in order to predict a continuous BP waveform. As such, the only available data set meeting those requirements is [13]. The results of the analysis using a randomly selected subset from [13] show that while the final results are similar, there is still some deviation in them, with the SBP MAE increasing from 2.7 to 6 mmHg and DBP MAE from 1.6 to 3.4 mmHg. This evidences that even when using the same data set and the same model there can be variations in the results. This may be due to different parameters used in the training process, for example. The model was reproduced using 10 epochs and a batch size of 512, but these parameters were not disclosed in the original paper, thus hindering the reproduction process.

Ibtehaz et al. [20]

This model was built using PPG data and outputs a continuous BP waveform; therefore, it could also be evaluated on the Aguirre et al. [16] data set. One of the main differences between the two data sets is their size. For example, 353.5 hours of signals are used as input in [20],

whereas only 48 hours of data are available in [16] when reproducing the paper. Figure 5 shows the vast improvements of the model when a larger data set is available illustrating the model performance when 4.8 hours, 48 hours, and 353.5 hours of data from [13] are adopted, respectively. It can be seen that 4.8 hours is not nearly a large enough training data set size to yield results close to resembling a BP signal. Even with 48 hours of data available DBP and SBP predictions substantially differ from those of the original paper. Table IV shows the percentage accuracy of each prediction using data from [16], relative to the original results. It is evident that less than half of SBP predictions are within 15 mmHg. However, based on Figure 5, it could be assumed that, given the complex architecture adopted, if a larger data set were available the accuracy could be comparable to the original paper. Nevertheless, the data set used in the original paper presented longer signals from a much reduced number of subjects compared to [16], thus it is possible that the accurate predictions reported in [20] were due to less variability present overall across the BP signals. Another issue with this model is the computing time, as model training required 12 days with the computational resources available when using 48 hs of data.

TABLE III. CUMULATIVE ERROR

		Cumulative Error Percentage (%)		
		≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg
Ibtehaz et al. [20]	DBP	82.83	92.16	95.73
	SBP	70.81	85.3	90.92
	Mean arterial pressure (MAP)	87.38	95.17	97.73
Reproduction of Ibtehaz et al. [20]	DBP	31.1	52.6	68.3
	SBP	17.8	34.1	49.4
	MAP	37.1	66.6	83.2

Li et al. [21]

It is clear that the results obtained from the different data sets are not as accurate as the ones obtained in the original paper. The model may have been over-fit in the original study since the accuracy of the model decreases when being trained/tested on a much larger number of cycles (6,852 vs 678,202 beats) with MAE results being 4-to-9 times worse. This can be due to the much more pronounced variability present in the signals when testing on a larger data set. Results on data sets [14] and [17] show even more pronounced decrements in performance. Again, this could be explained by the fact that the Kachuee et al. data set [13] is a prepared subset of the MIMIC database. Recordings that were intermittent or corrupted by motion were removed by the curators, which hinders the development of more generalizable models. Performance differences are amplified when dealing with non-ideal signals, which can be even more problematic when dealing with a LSTM model in which features extracted from each cycle are employed as sequential data, thus corrupted features due to non-ideal cycles may affect the time series and the related learning process. Finally, another limitation is the lower number of beats adopted in [14] and [17] for reproducing the model which may have contributed to the worse results.

Athaya et al. [22]

Results obtained are promising considering that for both data sets evaluated, the obtained MAE is lower than 8 mmHg and 5 mmHg, for SBP and DBP respectively. Data from 100 subjects were used in both reproductions. While recordings from the MIMIC data set were typically between 24 and 48 hours (with an average of 3.4 hours of data for each subject adopted into the original model), the length of the recordings in in [13] and [16] were much shorter. Based on the improvement in performance when using [13] compared to [16], and considering that data set [13] has longer recordings, it is possible to assume that longer recordings may be a reason for the improved accuracy.

Signal Quality

Based on the results shown in this analysis, it was assumed that differences in signal quality across the various data sets may be a factor in the deviance of performance. The Perfusion Signal Quality Index (PSQI), which is considered the gold-standard for PPG signal quality [25], was calculated for all data sets and is summarized in Table I. It is evident that no clear correlation is present between signal quality and the accuracy of predictions as shown in Table III. It would be expected that data set [17] would always provide the worst performance for each model in this case but this does not occur, indicating that signal quality is not an overwhelming factor in causing the lack of reproducibility.

V. DISCUSSION

While a significant number of studies have been produced in the last few years in the field of cuffless BP monitoring [26-28] based on ML/DL models, the external validation of those models is often an overlooked area which is, however, fundamental for the acceptability and pervasive adoption of the systems based upon them.

As discussed in [10], there is a substantial variation in the approaches adopted to build these systems, in the approaches adopted and, most importantly, in the protocols adopted for data collection. Even though ESH or AAMI protocols could be adopted for the validation of the developed systems, there is no standard clinical protocol available for data collection in those settings, with different groups of researchers adopting different types of measurement devices (for ECG and/or PPG, but also for gold-standard BP measures), number of subjects, eligibility and exclusion recruitment criteria (i.e., particular subjects' conditions, ethnicity), data cleaning and pre/post-processing procedures, settings (e.g., hospitals vs. labs), and duration of data recordings for each subject. Considering all of these issues, it becomes problematic to find equivalent data sets and the above factors contribute to the differences in performance shown in Section IV when reproducing models on multiple independent data sets; however, taking into account all these factors would allow for the development of a truly generalizable model for cuffless BP monitoring.

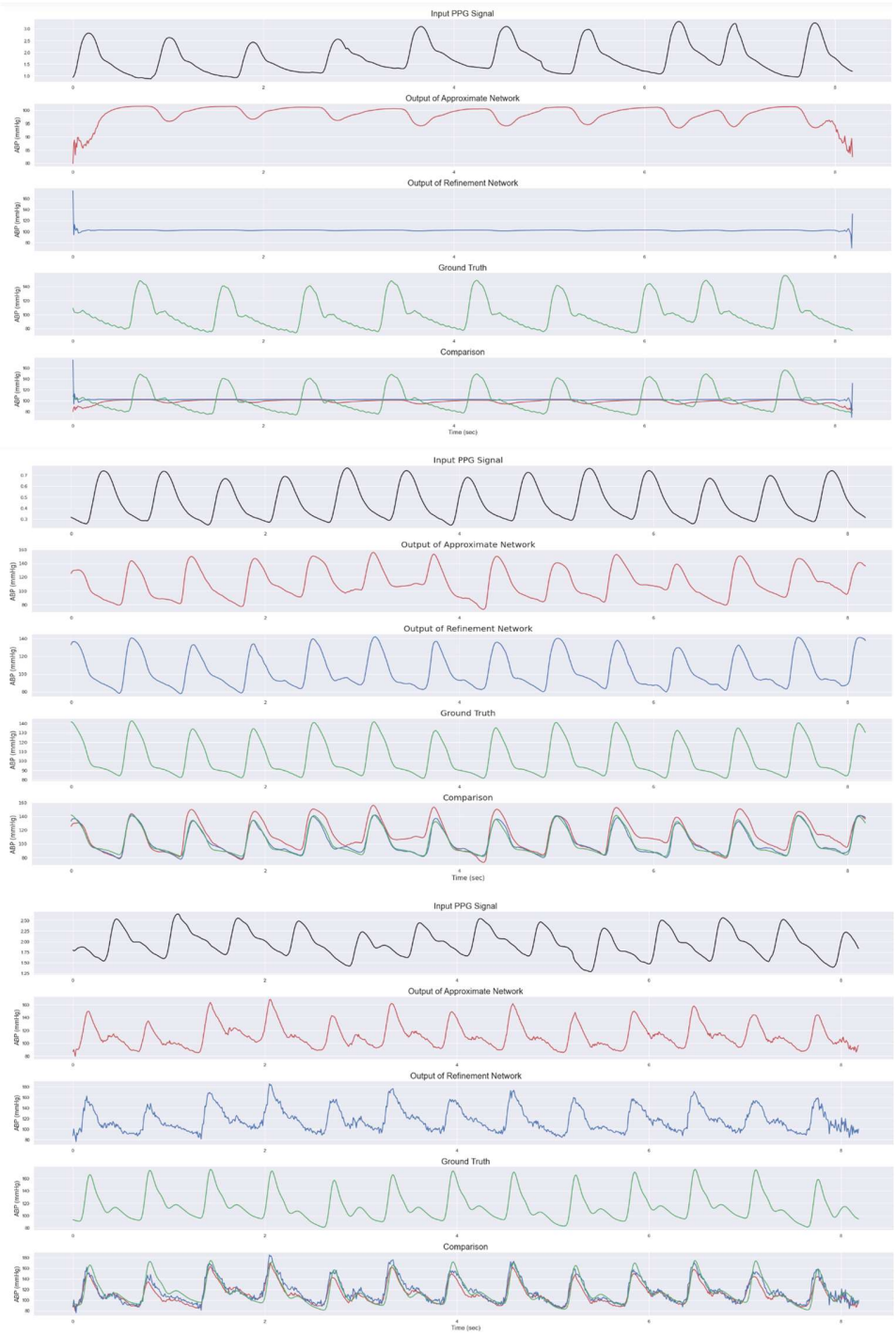

Figure 5. Model performance when 4.8 hours, 48 hours, and 353.5 hours of data from [13] are adopted. It is evident the difference in performance based on the amount of data used for training

For instance, due to BP variations between individuals tending to be much larger than within an individual over a short period of time, data sets with a lower number of subjects but longer recordings per subjects seemed to exhibit improved results; however, how well models built in such conditions could generalize over to the general population is still an open question. Moreover, feature extraction methods may not always be sufficiently robust or comprehensive to be adapted on other data sets. In general, based on results reproduced from [19] and [22], it can be assumed that using as input the raw time-series ECG and/or PPG signal in to a model might provide more robust predictions. Thus, longer signals and models built on the collected time-series are preferable for higher degrees of precision; also, U-Net architecture models with convolutional layers have provided promising results.

Another possible source of concern is the adoption of data sets from a very limited pool when building a model; for instance all models considered in this study were originally developed based on MIMIC and/or Kauchee et al. [13] data. This could be a potential problem as those data sets represent the standard benchmarks in the field but fail in adequately providing an overall reflection of the real-world, with the consequence that models built only using these data sets may not generalize well in real-world settings, especially for underrepresented demographic groups. Therefore, high-quality open access data, ideally based on standardized data collection protocols, are essential in the development of reproducible ML/DL models. In general, it has been reported that health-related ML models perform particularly poorly on reproducibility measures relative to other disciplines [29]; however, the curation of large open data sets covering institutions, countries, and populations, and the adoption of standards for collecting data and reporting ML studies would be a huge steps forward in the resolution of this challenge [29].

This present study has a number of strengths and limitations. To the best of the authors' knowledge, this is the first study to evaluate several of the currently most well cited ML/DL-based models for cuffless BP monitoring over multiple independent data sets. The results of this investigation have produced particular recommendations regarding standardized data collection protocols, models and signals, data recording length, and open access data as potential steps to overcoming the challenge of reproducibility in ML/DL models in this field and the health domain in general. Despite the best intentions, a complete reproduction of a model may be challenging due to a number of parameters used for model training being not disclosed by the authors in the original publications. It is thus evident that, together with open access data, also making the code developed by the authors publicly available would support the implementation of those reproducibility investigations in the health-related ML models. Moreover, considering data sets of such variations in terms of size allowed us to evaluate the several challenges present in adopting currently available datasets in BP monitoring, even though this may have contributed to the deviation in the results from the original studies as described in Section IV.

The variations in the considered data sets in terms of BP targets (waveform vs on-off measurements), signals adopted as input (e.g., ECG vs PPG vs both), etc. prevented us from using all data for every single model and from combining multiple datasets in this evaluation. A step forward in our analysis would be to adopt the most promising model architecture resulting from this analysis and train it on a subset of combined data set with similar characteristics and then test it using test data from each of the datasets of interest so as to evaluate its generalisation capabilities. However, the challenge of finding and developing equivalent data sets in order to tackle the limitations addressed above must be first overcome.

VI. Conclusion

Despite a plethora of works having been produced in this field of cuffless BP monitoring based on ML/DL, very limited research has been carried out in terms of reproducibility. To the best of the authors' knowledge, this is the first study to evaluate several of the currently most well cited ML/DL-based models for cuffless BP monitoring over multiple independent data sets. Results obtained with the present investigations showed that most of the models taken into account have significantly variable performance when adopted on different data sets and, sometimes, even when adopted on the original data used for building the model itself. Particular recommendations regarding standardized data collection protocols, models and signals, data recording length, open access data and code and possible future steps have been provided as potential steps to overcoming the challenge of reproducibility in this field.

References

[1] P.K. Whelton, "Epidemiology of hypertension," Lancet, vol. 344, no. 8915, pp. 101-106, 1994

[2] S. Romagnoli, et al., "Accuracy of invasive arterial pressure monitoring in cardiovascular patients: an observational study," Crit Care, vol. 18, 644, 2014

[3] E. O'Brien, et al., "Blood pressure measuring devices: Recommendations of the European Society of Hypertension," BMJ, 322, 531-536, 2001

[4] A.S. Meidert, B. Saugel, "Techniques for Non-Invasive Monitoring of Arterial Blood Pressure," Front Med., vol. 4, 231, 2018

[5] A. Patzak, et al., "Continuous blood pressure measurement using the pulse transit time: Comparison to intra-arterial measurement," Blood Press., 24, 4, 217-221, 2015

[6] S.N.A. Ismail, et al., "Recent advances in non-invasive blood pressure monitoring and prediction using a machine learning approach," Sensors, 22, 16, 6195, 2022

[7] N.D. Agham, U.M. Chaskar, "Learning and non-learning algorithms for cuffless blood pressure measurement: a review," Med Biol Eng Comput., 59, 6, 1201-1222, 2021

[8] M.M.R. Khan Mamun, A. Sherif, "Advancement in the Cuffless and Noninvasive Measurement of Blood Pressure: A Review of the Literature and Open Challenges," Bioengineering, 10, 1, 27, 2023

[9] G.S. Stergiou, et al. "Cuffless blood pressure measuring devices: review and statement by the European Society of Hypertension Working Group on Blood Pressure Monitoring and Cardiovascular Variability," Journal of Hypertension, 40, 8, 1449-1460, 2022

[10] S.M.S. Islam, et al., "Wearable cuffless blood pressure monitoring devices: a systematic review and meta-analysis," Eur Heart J - Dig Health, 3, 2, 323–337, 2022

[11] Y. Liang, et al., "Photoplethysmography and Deep Learning: Enhancing Hypertension Risk Stratification," Biosensors, 8,101, 2018

[12] B. Huang, et al., "MLP-BP: A novel framework for cuffless blood pressure measurement with PPG and ECG signals based on MLP-Mixer neural networks," Biomed Sign Proc Contr, 73, 103404, 2022

[13] M. Kachuee, et al., "Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring," IEEE Trans Biomed Eng, 64, 4, 859-869, 2017

[14] https://outbox.eait.uq.edu.au/uqdliu3/uqvitalsignsdataset/index.html

[15] Y. Liang, et al., "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China," Sci Data., 27, 5, 2018

[16] N. Aguirre, et al., "Blood Pressure Morphology Assessment from Photoplethysmogram and Demographic Information Using Deep Learning with Attention Mechanism," 2021 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4598938

[17] P. Mehrgardt, et al., “Pulse Transit Time PPG Dataset,” (version 1.1.0). PhysioNet. https://doi.org/10.13026/jpan-6n92.

[18] Y. Kurylyak, F. Lamonaca, D. Grimaldi, “A Neural Network-based method for continuous blood pressure estimation from a PPG signal,” 2013 IEEE Int Instrum. Meas. Technol. Conf., 280-283, 2013

[19] F.P.W. Lo, et al., “Continuous systolic and diastolic blood pressure estimation utilizing long short-term memory network,” IEEE Int Conf Eng Med Biol Soc (EMBC), 2017, 1853-1856

[20] N. Ibtehaz, et al., “PPG2ABP: Translating Photoplethysmogram (PPG) Signals to Arterial Blood Pressure (ABP) Waveforms,” Bioengineering, 9, 692, 2022

[21] Y.-H. Li, et al., “Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model,” Sensors, 20, 5606, 2020

[22] T. Athaya, S. Choi, “An Estimation Method of Continuous Non-Invasive Arterial Blood Pressure Waveform Using Photoplethysmography: A U-Net Architecture-Based Approach,” Sensors, 21, 1867, 2021

[23] G.D. Clifford, D.J. Scott, M. Villarroel, “User Guide and Documentation for the Mimic ii Database,” MIMIC-II Database Version 2009, Version 2.6. Available online: https://physionet.org/mimic2/UserGuide/UserGuide.pdf

[24] R. Mukkamala, et al., “Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice,” IEEE Trans Biomed Eng., 62, 8, 1879-901, 2015

[25] M. Cannesson, et al. “Does the pleth variability index indicate the respiratory-induced variation in the plethysmogram and arterial pressure waveforms?,” Anesth. Analg., 106, 1189–1194, 2008

[26] C. Crowe, et al. “Cuff-less blood pressure monitoring in a cohort of people with Parkinson’s disease,” submitted

[27] A. Valerio, et al. “Development of a PPG-based hardware and software system deployable on elbow and thumb for real-time estimation of pulse transit time,” submitted

[28] Z. Haider Janjua, et al., “Knowledge-driven feature engineering to detect multiple symptoms using ambulatory blood pressure monitoring data,” Comp Meth Prog Biom, 217, 2022

[29] E. Sohn, “The reproducibility issues that haunt health-care AI,” Nature, vol. 613, no. 7943, pp. 402-403, 2023