Artificial Intelligence Review
https://doi.org/10.1007/s10462-022-10353-8

Machine learning and deep learning for blood pressure prediction: a methodological review from multiple perspectives

Keke Qin[1] · Wu Huang[2] · Tao Zhang[1] · Shiqi Tang[3]

Abstract
Blood pressure (BP) estimation is one of the most popular and long-standing topics in health-care monitoring area. The utilization of machine learning (ML) and deep learning (DL) for BP prediction has made remarkable progress recently along with the development of ML and especially DL technologies, and the release of large-scale available datasets. In this survey, we present a comprehensive, systematic review about the recent advance of ML and DL for BP prediction. To start with, we systematically sort out the current progress from four perspectives. Then, we summarized commonly-used datasets, evaluation metrics as well as evaluation procedures (especially the usually ignored *splitting strategy* operation), which is followed by a critical analysis about the reported results. Next, we discussed several practical issues as well as newly-emerging techniques appeared in the research community of BP prediction. Also, we introduced the potential application of several advanced ML technologies in BP estimation. Last, we discussed the question of *what a good BP estimator should look like?*, and then a general proposal for an objective evaluation of model performance is given from the perspective of an ML researcher. Through this survey, we wish to provide a comprehensive, systematic, up-to-date (to Feb, 2022) review of related research on BP prediction using ML & DL methods, which may be helpful to researchers in this area. We also appeal an objective view of the progress reported in the relevant literatures in a more systematic manner. The experimental data & code and other useful resources are available at https://github.com/v3551G/BP-prediction-survey.

Keywords Blood pressure prediction · Machine learning · Deep learning · Multi-view taxonomy system · Physiological signal

1 Introduction

Background Blood pressure (BP) is an important dynamic physiological index reflecting personal health status, which is often used for health monitoring and disease prevention (MacMahon et al. 1990; Singla et al. 2019). Systolic BP (SBP) and diastolic BP (DBP) are

✉ Wu Huang
huangwu@scu.edu.cn

Extended author information available on the last page of the article

Published online: 26 December 2022

two crucial indicators of BP. BP monitoring is an extensively studied topic in healthcare monitoring area. In fact, the study of BP prediction can be traced back to over 100 years ago (Buchanan et al. 2011). From the earliest mercury sphygmomanometers to the latter oscillometric method and auscultation method, etc., these methods are all physical methods based on pressure and can not be used for continuous BP monitoring. Pulse transit time (PTT) methods (Mukkamala et al. 2015; Peter et al. 2014; Sola et al. 2013) can be used for continuous BP monitoring. It is, however, an ideal (linear) model, and the frequent calibration over time has to be performed on an individual basis in order to ensure accuracy (Ding et al. 2017; Samartkit et al. 2022).

Thanks to the advances of machine learning (ML) and deep learning (DL) technology and the release of several large databases that are freely accessibly, ML and DL has come into the spotlight as a very useful, non-invasive approach for BP prediction by using biosensors. This is clearly reflected in the corresponding growth of relevant publications, as Fig. 1a illustrates. This type of methods are inherently data-driven where prediction model is trained using ML and DL with the aid of large amount of training data, which actually leverages the powerful capabilities of DL in feature learning, expression and modeling of complex relationships (LeCun et al. 2015).

Motivation Although there have been several high-quality surveys about BP estimation (Maqsood et al. 2022; Mukkamala et al. 2021; Picone et al. 2017; Drawz et al. 2012; Hosanee et al. 2020; Chao et al. 2021; El-Hajj and Kyriacou 2020b; Martinez-Ríos et al. 2021; Forouzanfar et al. 2015; Tamura 2021; Steinman et al. 2021), they mainly focus on traditional methods (refer Table 19). ML methods, especially DL methods, are rarely or not adequately covered, because the application of DL in physiological signals is relatively lagging behind, and most of them are mainly published during and before the outbreak of DL. Therefore, it is necessary to systematically sort out the latest progress in this area. Second, in the era of DL, there are some new emerging issues (such as the comparison between hand-crafted features and machine-learned features, etc.) and techniques (such as data augmentation, signal combination scheme, etc.) related to BP estimation worth discussing. Moreover, based on our review of over 200 papers on BP prediction that have been published in various journals and conferences (as Fig. 1b depicts), we found some key but neglected factors related to the problem of reproducibility as well as the objective evaluation of model's performance, from the perspective of an ML researcher. Therefore,

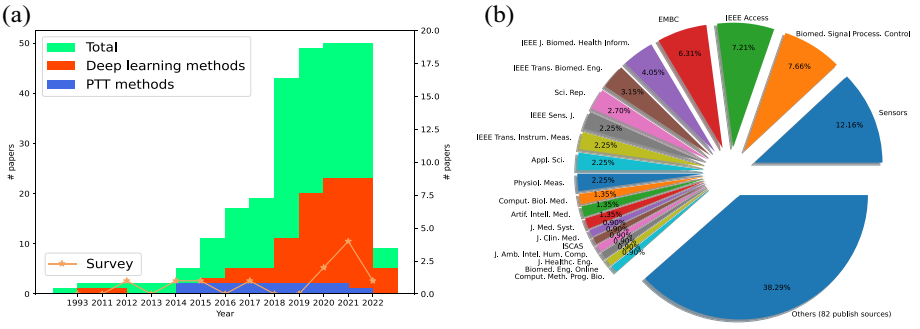

Fig. 1 The publication trend and distributions of publication sources of the literatures for blood pressure estimation mainly from 2011 to 2022. **a** the publication trend of papers based on machine learning, especially deep learning, and review papers on blood pressure prediction; **b** the distribution of publication sources (including a total of about 100 journals or conferences)

we plan to conduct a thorough review and analysis of the latest progress in all aspects of data-driven BP estimation.

Differences from existing surveys The main differences between our study and the existing ones are summarized as follows: (1) *systematic, comprehensive review.* Without being limited to a specific signal source, measurement method and measurement scene, we provide a review of blood pressure prediction community that has the following characteristics, (i) systematic: current progresses are sort out based on the proposed multi-aspects taxonomy (as illustrated in Fig. 2), (ii) comprehensive: all elements of the construction of blood pressure prediction pipeline are involved. (2) *more recently published works.* The publications in the last four years (2018–2021) are much more than all those published before 2018 (as illustrated in Fig. 1a). In the light of the evolution of ML and DL technologies and its widely application in BP prediction area in the past few years, this survey covers extensively the recent published studies. Therefore, we provide the up-to-date reviews of the newly presented methods. (3) *critical thinking and a general proposal.* We critically thought the unfairness of system comparisons from a machine learning perspective, and analyzed the factors that lead to the unreliability of results reported in related studies from multiple aspects, so as to propose a general proposal towards objective assessment of model's performance, and put forward suggestions for the future research directions.

Specifically, there are several surveys (Hosanee et al. 2020; El-Hajj and Kyriacou 2020b; Martinez-Ríos et al. 2021) about BP prediction, most of which, however, only focus on traditional methods such as pulse transit time (PTT), pulse wave velocity (PWV), pulse arrival time (PAT), pulse wave analysis (PWA), and traditional feature-based ML methods, especially DL methods are not or rarely covered. Besides, another set of surveys only focused on certain aspects (e.g the limitations of conventional evaluation standards and analyzing tools (Mukkamala et al. 2021), the accuracy of cuff-based BP (Picone et al. 2017), the cuffless BP monitor standards and approval for medical use (Tamura 2021) and the usage scenario (Drawz et al. 2012) of BP prediction pipeline. In addition, certain surveys are limited to specific signal source such as Photoplethysmography (PPG) signal (Hosanee et al. 2020; El-Hajj and Kyriacou 2020b; Maqsood et al. 2022), oscillometric waveform (Forouzanfar et al. 2015) or facial video (Steinman et al. 2021). The only two surveys that DL methods are covered is written by Chao et al. (2021), and Maqsood et al. (2022), respectively. In these surveys, related work are grouped into traditional feature-based ML methods and DL methods according to

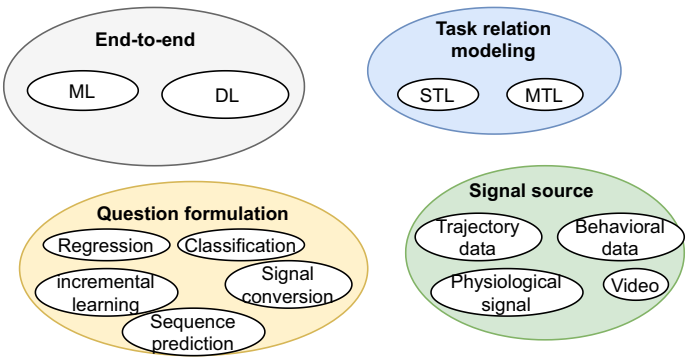

Fig. 2 The schematic diagram of a multi-view classification system for work related to blood pressure prediction

whether there is explicit feature extraction. However, the classification of DL methods is not comprehensive enough and the granularity is too coarse. Moreover, a considerable number of newly emerging studies were not included.

Instead of limited to specific signal source or specific measurement method, this survey focuses on systematically and comprehensively categorizing and reviewing the latest progress of ML and DL for BP prediction. Specifically, we reviewed the current progress (to Feb, 2022) of BP prediction from a total of four aspects, especially those DL methods. In addition, several practical issues/technologies involved in the whole research pipeline are discussed/summarized in detail, as well as the potential application of several advanced ML topics in this field.

Concepts We found that some basic concepts were overlooked or confused in relevant studies. Therefore, the definition of these concepts are firstly announced in Table 1 to avoid ambiguity. In addition, suppose each individual/subject contains only a record of data, and therefore 'Record', 'Individual', and 'Subject' have similar meanings, provided that there is no ambiguity. Besides, for ease of reading, all abbreviations appearing in the paper are summarized in Table 23.

Contributions Summarily, the main contributions of this survey include:

- We reviewed all the elements required for the construction of BP prediction's pipeline: datasets and processing tools, data preparation (includes signal denoising, data cleaning, feature engineering, and feature selection/reduction), training algorithms, evaluation metrics, and evaluation strategies, etc (Sects. 3, 4).
- We provide a comprehensive survey of the utilization of machine learning and deep learning for blood pressure prediction. Specifically, we build a multi-aspect taxonomy to present elaborated categorizations of current advances of blood pressure research, in an attempt to make the readers understand them in a systematic way (Sect. 3).
- We systematically reviewed some critical while practical issues/techniques (such as imbalance phenomenon, sample duration, data augmentation, individual difference, signal combination schemes, etc.) in blood pressure estimation area and introduced several potential, advanced machine learning topics (such as Auto ML, transfer learning, meta learning, federated learning, etc.) (Sects. 5, 6).

Table 1 Some basic notations

Concept	Explanation
Record	A record is a collection of an individual's physiological signal (ABP signal included) over a period of time
Segment	Sampling point sequence. For example, a segment of T seconds of a signal with sampling frequency of fs contains $T \cdot fs$ sampling points.
Sample	The basic unit of a dataset in machine learning. A sample includes an input vector (a signal segment or a feature vector extracted from the signal segment) and the ground-truth target (BP value or ABP segment)
Dataset	A collection of samples. A standard dataset usually includes training set, validation set (optional), and test set, which is used for model training, model choice/validation, and model test, respectively
I.I.D assumption (Bishop and Nasrabadi 2006)	Independent-identical-distribution (I.I.D), i.e all samples are sampled independently from an 'unknown' distribution, and the distribution of training data and test data should be the same, which is an assumption in conventional machine learning realm

- We analyzed the unfairness of systematic comparison that widely used in this area and disclosed the factors leading to unreliability of the results reported in related studies from multiple aspects (data preparation, feature selection, normalization, evaluation metrics, evaluation strategies, etc.), which results in a final proposal towards the objective evaluation of blood pressure prediction model (Sects. 4, 7).

Organization The rest of this survey is organized as follows. Section 2 provides a brief review of the explicit analytical models and data-driven BP predictions, respectively. Section 3 systematically reviews the current progress of BP prediction from four dimensions. Section 4 introduced some widely used datasets, evaluation strategies in this area, and our critical analysis of the current progress in data-driven BP estimation. Section 5 investigates some practical issues as well as newly-emerging techniques in this area. Section 6 discussed the application of some advanced machine learning topics in this area. Finally, Sect. 7 gives our general discussion of the BP prediction problem as well as conclusions and future research directions. A more detailed overall schematic diagram of this survey is presented in Fig. 3.

2 From explicit analytical model to data-driven BP prediction

Explicit analytical model The most well-known analytical method for non-invasive BP estimation is PTT/PAT/PWV. The basic physics behind such methods is arterial wall mechanisms and wave propagation in the arteries (Mukkamala et al. 2015), where the former builds the relationship between BP and arterial elasticity through Hughes equation, and the latter establishes the relationship between arterial elasticity and PTT or PWV through Moens-Korteweg (MK) equation or Bramwell-Hill (BH) equation (Samartkit et al. 2022;

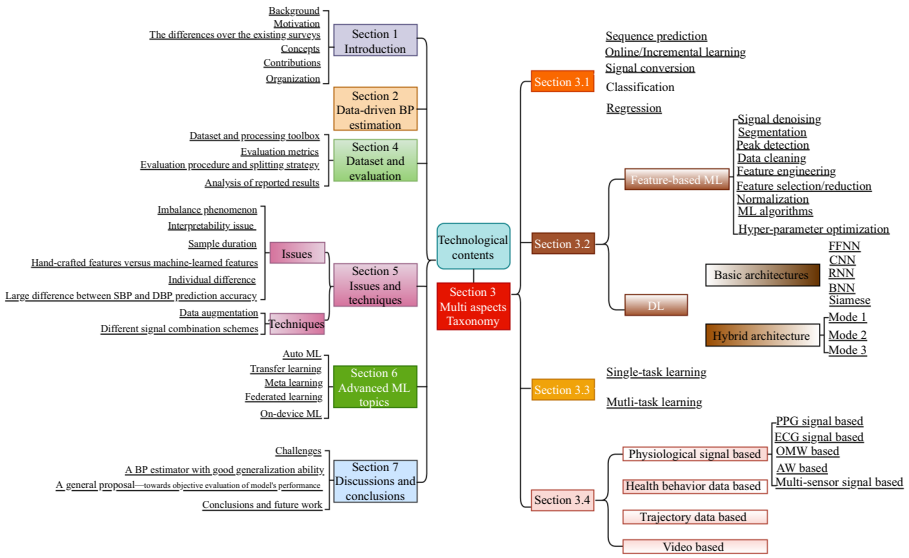

Fig. 3 The structure of this survey. The left part (Sects. 1, 2, 4, 5, 6, 7) refers to the functional contents that provide overall introduction, review and discussion. The right part (Sect. 3) refers to the technological contents that provide detailed reviewing of related work from four perspectives

Chen et al. 2000). Finally, the relation between BP and PTT or PWV is established. The MK equation is established on the assumption that the artery wall can be modeled as a thin shell, and the thickness and radius of the artery remain fixed as the BP changes (Ma et al. 2018). These assumptions, however, may not hold for human arteries. Ding et al. (2015, 2017) extends the classical PTT method by introducing a new arterial diameter change indicator-PIR (Ding and Zhang 2015) to capture the low-frequency components of BP which originates from peripheral resistance. In order to consider the neglected non-Newtonian fluid properties of blood, Thambiraj et al. (2019) further extends Ding's work by introducing a viscous flow indicator-Womersley number. Recently, Ma et al. (2018) proposed a new analytical model that correlates BP with PWV without the above assumptions mentioned and empirical Hughes equation. In addition, Matsumura et al. (2018) directly correlates BP with cardiac output and total peripheral resistance, which were estimated with heart rate and modifed normalized pulse volume, respectively. Table 2 summarizes the analytical models of the above-mentioned work for intuitive comparison.

In general, despite having intuitive and easily interpretable mathematical expressions, the models as mentioned above with limited expressive power that dependents on only a few factors are based on certain ideal assumptions that may not hold in practice. Besides, there are several challenges for implementing PTT-based BP monitoring. First, the biggest challenge is the need for calibration (Mukkamala et al. 2015). Due to individual differences and dynamic cardiovascular changes over time, the parameters involved are all subject-specific and has to be calibrated over time on an individual basis (Ding et al. 2017; Samartkit et al. 2022). Therefore, PTT method is usually employed for individualized BP estimation. Second, a practical issue is the convenient measurement of at least two waveforms at different sites for robust estimation of PTT/PWV. Therefore, the configuration complexity and power load of the sensor, the convenience and stability when wearing and the quality of the collected signal must be considered (Samartkit et al. 2022; Mukkamala et al. 2015; Ding et al. 2017). Third, the need for determination of SBP, DBP, and mean BP (MBP), independently (Sharifi et al. 2019; Mukkamala et al. 2015), since these three BP measurements are of clinical importance. However, due to the existence of isolated systolic hypertension that usually occurs in the elderly, conventional PTT correlates less well with SBP.

Data-driven BP prediction Different from conventional pressure-based physical methods or explicit analytical/mathematical methods (such as PTT) inspired by physiological mechanism or basic physics, the goal of *data-driven BP prediction* is to learn the unknown non-linear relationship (Monte-Moreno 2011) between input signal and BP using ML or DL technologies with the help of a large number of training data, in a supervised learning mode. Of course, the premise that the learned relationship is meaningful is that the two are strongly correlated. Data-driven methods actually provide a possible way to realize a general BP prediction model-only all relevant factors affecting BP need to be taken as inputs, thus avoiding frequent calibration. In data-driven approaches, data plays an extremely important role, and each link of data flow (including data collection, data cleaning, data labeling, data strategy of monitoring model, etc.) will affect the credibility of the final training model (Liang et al. 2022). Depending on the specific use scenario, there are many signal sources available for BP prediction, signal sources used for non-invasive, data-driven BP prediction include *physiological signal*, *health behavior data*, *trajectory data*, and *facial video*, which all carry important information related to BP changes. Table 3 summarizes several representative BP prediction methods using each type of data source.

For physiological signal, PPG and electrocardiograpshy (ECG) signal are the most popular signals used for BP prediction (Maqsood et al. 2022). Generally, PPG signal depicts the hemodynamics in the peripheral vasculature of the individual, which reflects the total peripheral

Table 2 Summary of several analytical models for non-invasive blood pressure estimation

References	Correlation	BP calculation formula (physical model)	Data required
Chen et al. (2000)	$BP \propto PTT$	$SBP = SBP_0 - \frac{2}{\gamma PTT_0}(PTT - PTT_0)$ $DBP = DBP_0 - \frac{2}{\gamma PTT_0}(PTT - PTT_0)$	PPG, ECG
Samartkit et al. (2022)	$BP \propto \mathbf{ln}\frac{1}{PTT}$	$SBP = \frac{2}{b}\mathbf{ln}\frac{a}{PTT_s}$ $DBP = \frac{2}{b}\mathbf{ln}\frac{a}{PTT_d}$	PPG, PZT
Ding et al. (2015)	$PP \propto \frac{1}{PTT^2}$ $DBP \propto \frac{1}{PIR}$	$SBP = DBP_0 \cdot \frac{PIR_0}{PIR} + PP_0 \cdot (\frac{PTT_0}{PTT})^2$ $DBP = DBP_0 \cdot \frac{PIR_0}{PIR}$	PPG, ECG
Ding et al. (2017)	$PP \propto \frac{PIR}{PTT^2}$	$SBP_i = MBP_0 \cdot \frac{PIR_0}{PIR_i} + \frac{2}{3}\frac{PIR_i}{PIR_0}(\frac{PTT_0}{PTT_i})^2$ $DBP_i = MBP_0 \cdot \frac{PIR_0}{PIR_i} - \frac{1}{3}\frac{PIR_i}{PIR_0}(\frac{PTT_0}{PTT_i})^2$	PPG, ECG
Thambiraj et al. (2019)	$PP \propto \frac{PIR}{PTT^2}(1 - 0.56\frac{1}{\sqrt{2}\alpha})^2$	$MBP = MBP_0 + \frac{2}{\gamma}[\frac{1}{2}\mathbf{ln}\frac{PIR}{PIR_0} + \frac{1}{2}\mathbf{ln}(\frac{PTT_0}{PTT})^2 + \frac{1}{2}\mathbf{ln}(\frac{1-0.56\frac{1}{\sqrt{2}\alpha}}{1-0.56\frac{1}{\sqrt{2}\alpha_0}})^2]$ $DBP = MBP_0 + \frac{2}{\gamma}[\frac{1}{2}\mathbf{ln}\frac{PIR}{PIR_0} + \frac{1}{2}\mathbf{ln}(\frac{PTT_0}{PTT})^2 + \frac{1}{2}\mathbf{ln}(\frac{1-0.56\frac{1}{\sqrt{2}\alpha}}{1-0.56\frac{1}{\sqrt{2}\alpha_0}})^2] - \frac{1}{3}PP$ $SBP = PP + DBP$	PPG, ECG
Ma et al. (2018)	$\frac{PWV}{\sqrt{\frac{C \cdot e^{a_2 E_{cc}^2}}{\rho}}} = f(\frac{P}{C \cdot e^{a_2 E_{cc}^2}}, a_1, \frac{h_0}{R_0})$	–	PPG, ECG
Matsumura et al. (2018)	$BP = CO \times TPR$	$MBP = \mathbf{exp}(a_{MBP} \cdot \mathbf{ln}HR + b_{MBP} \cdot \mathbf{ln}mNPV + c_{MBP})$ $SBP = \mathbf{exp}(a_{SBP} \cdot \mathbf{ln}HR + b_{SBP} \cdot \mathbf{ln}mNPV + c_{SBP})$ $DBP = \mathbf{exp}(a_{DBP} \cdot \mathbf{ln}HR + b_{DBP} \cdot \mathbf{ln}mNPV + c_{DBP})$	PPG

Table 3 Summary of representative methods using each data source for blood pressure estimation

Data source	Representative methods
Physiological signal	Kachuee et al. (2016), Monte-Moreno (2011), Chowdhury et al. (2020), Fan et al. (2019),Simjanoska et al. (2020), Miao et al. (2019), Miao et al. (2017), Slapničar et al. (2019), Leitner et al. (2021),Haddad et al. (2021), Baek et al. (2019), Su et al. (2018), Schlesinger et al. (2020), Ji et al. (2022)
Health behavior data	Chiang and Dey (2018) and Chiang et al. (2021)
Trajectory data	Xiang et al. (2021)
Facial video	Takahashi et al. (2020), Zhou et al. (2019), Luo et al. (2019), Djeldjli et al. (2021), Rong and Li (2021b),Schrumpf et al. (2021a)

vascular resistance (TPR) and cardiac output (CO) that are closely related to BP. Monte-Moreno (2011) first succeeded in building a machine learning system for BP prediction using extensive features derived from PPG signal inspired by the strong relation between physiological factors and BP. Lin et al. (2020) investigated the physiological mechanism of PPG for BP prediction based on feature analysis, and finds that each examined feature was TPR and/or CO correlated. ECG signal represents the electrical activity during heart function, and also contains BP-related information (Wu et al. 2016). Attia et al. (2019) found that ECG signals can be used to assess the cardiac contractility, which is one of the critical factors leading to the changes of BP.

Health behavior data (e.g. exercise, sleep, smoking, alcohol use, etc.) has been widely acknowledged as closely related to human health condition (Chiang and Dey 2018), and also further related to BP since BP is one of the most significant indicators of human condition. For example, Cornelissen and Smart (2013) has confirmed that exercise is statistically correlated with BP. Phillips et al. (2022) confirmed the direct nature of the association of alcohol use with BP. These factors actually act as mediators to influence BP, which in turn can be used to regulate BP in an active intervention circumstance.

For trajectory data, individual's daily routine inferred from trajectory to a certain extent reflects the regularity of routine, working pattern, and stress level, etc., all of which are closely related to BP level (Pickering et al. 1982).

For facial video, in every cardiac cycle, due to cardiac ejection, the collected facial video contains information of hemoglobin concentration changes over time. Blood flow pulsation in the cardiovascular can therefore be detected by capitalizing on subtle changes in skin color from the difference in re-emitted light between hemoglobin and melanin chromophores (Luo et al. 2019), based on computer vision (CV) technologies. The blood flow pulsation information can be further used to build BP prediction models.

3 Multi aspects taxonomy of BP prediction methods

As an application field of ML/DL, BP estimation usually includes the estimation of SBP, DBP and MBP. Blood pressure prediction is not limited to a single learning scheme, and all kinds of ML/DL technologies has been applied to BP prediction in related studies. At

the same time, there are many signal sources for BP estimation. Based on the above considerations, we try to sort out the relevant work from the following four perspectives:

(1) Taxonomy 1-how to model the question of BP prediction from the perspective of machine learning? From this perspective, related work can be divided into five categories, please refer Sect. 3.1;
(2) Taxonomy 2-whether feature extraction and predictive model building are performed simultaneously? From this perspective, related work can be divided into two categories, please refer Sect. 3.2;
(3) Taxonomy 3-whether the relationship among different tasks is modeled? From this perspective, related work can be divided into two categories, please refer Sect. 3.3;
(4) Taxonomy 4-the signal source used for building predictive model. From this perspective, related work can be divided into four categories, please refer Sect. 3.4.

3.1 Taxonomy 1: question formulation

Preliminaries In classical ML settings, the I.I.D assumption is followed and the entire training data is required to be made available prior to the learning task. Theoretically, a predictive model is determined by minimizing the expected risk as follows,

$$f^* = \arg\min_{f \in F} \mathbf{E}_{(x,y)\sim D} L(f(x;\theta), y), \tag{1}$$

where $L(,)$ denotes loss function, D denotes the unknown genuine distribution that generating sample (x, y), F denotes the assumption space, $f \in F$ is parameterized by θ. However, since the distribution D is usually unknown, in practice, a model is determined by minimizing the empirical risk (ERM). In addition, to overcome overfitting issues, an additional regularization term is used to control the model complexity, Therefore, in practice, a model is determined offline by minimizing structural risk (SRM) as follows,

$$f^* = \arg\min_{f \in F} \sum_{i=1}^{N} L(f(x_i;\theta), y_i) + \lambda J(\theta), \tag{2}$$

where on the right side of the Eq. (2), the first term represents empirical risk, the second term represents structural risk and is weighted by parameter λ for trade-off between the two terms.

However, the I.I.D assumption is too strict. In real life, the collected data usually shows obvious temporal dependency and the I.I.D assumption may no longer be tenable, i.e $p(y_t|x_t, x_{t-1}, \ldots, x_{t-n}) \neq p(y_t|x_t)$. In this settings, additional mechanisms are needed to model this temporal dependency, although the model is still trained offline.

In addition, in many actual scenes, data arrives in a sequence manner, often accompanied by concept drift, which is ubiquitous in streaming environment (He et al. 2011). It is obvious that the I.I.D assumption is severely violated. In this settings, learning and decision-making are carried out alternately. In general, after T rounds are passed, the goal of an online learner (Hoi et al. 2021) is to minimize the regret-R_T of the learner's predictions against the best fixed learner, which is defined as,

$$\mathrm{R}_T = \sum_{t=1}^{T} L(f(x_t;\theta), y_t) - \min_{\theta} \sum_{t=1}^{T} L(f(x_t;\theta), y_t), \tag{3}$$

Table 4 Comparison of different learning paradigms for blood pressure prediction

Learning paradigm	Online/offline	Temporal correlation	Output
Classification	Offline	No	BP category
Regression	Offline	No	BP value
Signal conversion	Offline	No	BP waveform
Sequence prediction	Offline	Yes	BP value
Online/incremental learning	Online	Yes	BP value

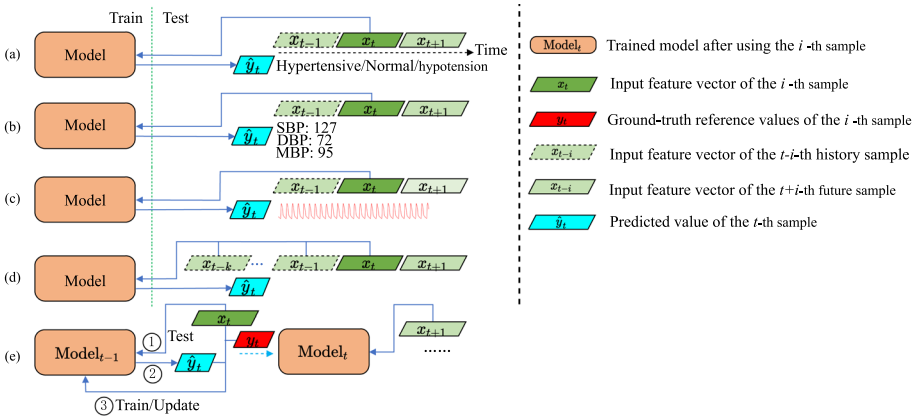

Fig. 4 Five formulations for blood pressure prediction. **a** classification question; **b** regression question; **c** signal conversion; **d** sequence prediction; **e** online/incremental learning

In this subsection, we focus on how to formulate BP prediction question from a machine learning perspective. We summarized a total of five learning scenarios widely used in the existing literatures for BP estimation, namely *classification question*, *regression question*, *signal conversion*, *sequence prediction*, and *online/incremental learning*. Figure 4 visually depicts the five paradigms of BP estimation, the main features of which are summarized in Table 4. The former three scenarios follow Eqs. (1) and (2). The fourth scenario still follows an optimization problem of similar form to Eq. (1), except that in addition to the current input, f is also conditioned on the previous input. The fifth scenario follows Eq. (3).

3.1.1 Classification question

In classification scenarios, the total BP range is divided into several disjoint intervals according to BP stages, each of which represents an independent category. Then BP monitoring is formulated as a two classes or multi-classes classification question, and a model is trained to predict the belonging category given input. In this paradigm, $L(,)$ usually means zero-one loss, logistic loss, softmax loss, etc. y denotes the index of classes. A few works are based on this paradigm. For example, Riaz et al. (2019) built an autoregressive-based ensemble model for identifying whether patients' BP is normal. El Attaoui et al. (2020) develops an embedded system combined with a wireless medical sensor network to detect the status of BP (normal or abnormal) in real time. Tjahjadi et al. (2020) developed

a bidirectional long short-term memory (Bi-LSTM) model for predicting the category of BP (normotension, prehypertension, and hypertension). Lee and Chang (2019) built a deep Boltzmann machine with Dempster-Shafer fusion to classify and estimate BP (10 categories) using oscillometric waveform.

Apparently, this type of modeling methods can only be used to diagnose the BP status (such as hypertension, hypotensive, normal, etc.) or predict the rough BP interval of an individual. In addition, the order of BP values is missing, which may lead to extremely abnormal results of the predictive model output.

3.1.2 Regression question

In regression scenarios, a model is directly trained to predict BP. In this paradigm, $L(,)$ usually means absolute loss or squared loss, y denotes ground-truth BP value. Almost all studies in this area are based on this paradigm owning to the continuous nature of BP value. In practice, due to the large range of possible BP values, *normalize target* technique is commonly used in DL-based methods (Song et al. 2021; Zhang et al. 2020b; Abrar et al. 2020; Athaya and Choi 2021; Aguirre et al. 2021; Song et al. 2019; Panwar et al. 2020; Mahmud et al. 2022; Tazarv and Levorato 2021) to boost gradient-based training.

3.1.3 Signal conversion

Recently, a few researchers have tried to predict BP indirectly by reconstructing ABP signal. Since BP value is parameter of ABP waveform, BP value can be acquired once high-quality ABP waveform is reconstructed. Signal conversion can be viewed as a generalized regression question. In this paradigm, $L(,)$ usually means absolute loss or squared loss, x and y denote the input signal and the target (i.e. ABP) signal fragment, respectively. As far as we know, Landry et al. (2019) firstly investigated the feasibility of generating ABP waveform using ECG signal. Ibtehaz and Rahman (2020) firstly attempted to translate PPG signal into ABP waveform using a deep learning model-U-Net. Athaya and Choi (2021) did similar thing as Ibtehaz and Rahman (2020). Sadrawi et al. (2020) built a deep convolution autoencoder model based on LetNet-5 and U-Net for PPG-to-ABP conversion. Cheng et al. (2021) built a U-Net based model for reconstructing ABP signal using PPG signal and its derivatives, and the maximum absolute loss is introduced in addition to squared loss to enforce the consistency of local characteristics between the predicted and the genuine ABP signal. Li and He (2021) built a generalized regression neural network model for single-period PPG-to-ABP conversion. Aguirre et al. (2021) built a Seq2Seq with attention model for PPG-to-ABP conversion. Harfiya et al. (2021) built an LSTM-based autoencoder model for PPG-to-ABP conversion. Qin et al. (2021) developed a convolution-based autoencoder model for PPG-to-ABP conversion, and domain adversarial training is introduced to conquer individual differences. There are also certain studies (Brophy et al. 2021; Mehrabadi et al. 2022) where the well-known CycleGAN was employed to learn the bijection between PPG signal and ABP waveform.

In addition to the above DL-based methods, Dash et al. (2020) proposed a subject-specific mathematical model based on the linear transfer function (LTF) technique for PPG-to-ABP conversion. Magbool et al. (2021) proposed a hybrid method that combine machine learning with the cross-relation blind estimation approach for reconstructing beat-by-beat ABP signal.

3.1.4 Sequence prediction

Different from the first three paradigms based on I.I.D hypothesis, sequence prediction enables modeling the underlying dependency between adjacent samples. Specifically, the current output y_t is related not only to the current input x_t, but also to historical data $x_{t'}, t' < t$. In this paradigm, both input x and output y are sequence. Models used for sequence prediction in this area includes recursive neural networks (e.g. Elman) (Wang et al. 2017) , recurrent neural networks (RNNs) (Senturk et al. 2020; Li et al. 2017; Tanveer and Hasan 2019; Su et al. 2018), nonlinear autoregressive model with exogenous input (NARX) (Senturk et al. 2020; Landry et al. 2019) and neural network output-error (NNOE) (Paviglianiti et al. 2020a, b), etc.

Popular RNNs used include the standard RNN and its variants such as LSTM, Bi-LSTM, GRU, etc. As Fig. 5a illustrates, in addition to the input x_t at current time step t, the current output also dependents on the hidden state of the previous time step. For RNNs, suppose $X_T = [x_1, x_2, \ldots, x_T]$ the input sequence, $Y_T = [y_1, y_2, \ldots, y_T]$ the target BP sequence. The conditional distribution $P(Y_T|X_T)$ is factorized as:

$$P(Y_T|X_T) = \prod_{t=1}^{T} p(y_t|h_t), \tag{4}$$

where hidden state h_t models the BP dynamics, h_t is generated from current input x_t and previous hidden state h_{t-1} as follows:

$$h_t = f(x_t, h_{t-1}). \tag{5}$$

NARX is a kind of nonlinear autoregressive model with exogenous inputs. As Fig. 5c illustrates, NARX use previous genuine target value and exogenous inputs (e.g. PPG, ECG, etc.) to predict the next target value. Formally, the shape of the regression vector is expressed as:

$$h(t) = [\underbrace{y(t-1|\theta), \ldots, y(t-n|\theta)}_{\text{autoregressive}}, \underbrace{x(t-d), \ldots, x(t-d-m)}_{\text{exogenous input}}], \tag{6}$$

where n is the y-predicted lag, m is the input lag and d the delay to obtain the prediction. The prediction vector is formulated as:

$$\hat{y}(t|\theta) = f(h(t), \theta), \tag{7}$$

where function f is implemented by neural network. Note that the previous genuine target y in the regression vector is replaced with predicted value $\hat{y}$ in the test phase.

NNOE (Norgaard et al. 2000) is a kind of neural network that models nonlinear dynamic system in stochastic environment. As Fig. 5b illustrates, NNOE is similar to

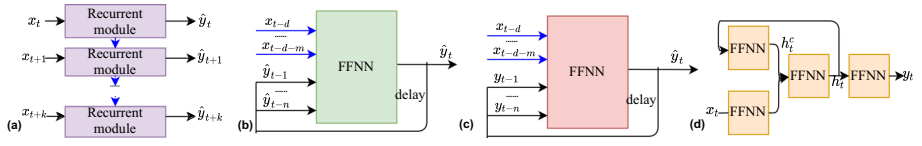

Fig. 5 Several classical model architectures for sequence prediction. **a** RNNs; **b** NNOE; **c** NARX; **d** Elman NN

NARX except that the genuine target value y in the regression vector is replaced with previous prediction value $\hat{y}$. Related work includes (Paviglianiti et al. 2020a, b), etc.

Elman is a kind of neural network model with local feedback, as Fig. 5d illustrates. Specifically, the addition contextual layer can remember the output of the hidden layer before the current time step, which enables Elman the ability of modeling time-related features. Formally, the hidden layer output is computed as $h_t = f(h_t^c, x_t)$, where $h_t^c = \alpha \cdot h_{t-1}^c + h_c$, and the final output is computed as $y_t = g(h_t)$. Related work includes (Wang et al. 2017), etc.

In addition, Sharifi et al. (2019) proposed a dynamic method based on the reconstruction of the state space of the cardiopulmonary system for BP prediction, where both current state and the past dynamical state based on state space reconstruction are jointly used for prediction. Formally,

$$\begin{cases} y_{\mathrm{DBP}}(n) = f_1(\mathrm{PIR}_n) + f_2(\mathrm{PIR}_{n-\tau}, \mathrm{PIR}_{n-2\tau}, \ldots, \mathrm{PIR}_{n-m\tau}), \\ \mathrm{PP}_n = f_1(\mathrm{PTT}_n) + f_2(\mathrm{PTT}_{n-\tau}, \mathrm{PTT}_{n-2\tau}, \ldots, \mathrm{PTT}_{n-m\tau}), \\ y_{\mathrm{SBP}}(n) = y_{\mathrm{DBP}}(n) + \mathrm{PP}_n, \\ y_{\mathrm{MBP}}(n) = y_{\mathrm{DBP}}(n) + 0.01 \cdot \mathbf{exp}(4.14 - \frac{40.74}{\mathrm{HR}}) \cdot \mathrm{PP}_n, \end{cases} \tag{8}$$

where functions f_1 and f_2 are learned by the multi-adaptive regression spline (MARS) method.

3.1.5 Online/Incremental learning

Unlike the above four mentioned scenarios that the entire training data has to be made available in advance and the model training is performed in an offline manner, incremental/online learning (Hoi et al. 2021; He et al. 2011) is new learning technique that learn models incrementally from data in a streaming manner. Intuitively, in online learning, current model firstly tries to make decision when a new sample arrives, and then the sample is used to update the model in a supervised mode. In other words, the prediction and model update are performed alternatively.

Chiang and Dey (2019) firstly proposed a random forest with feature selection (RFFS) model coupled with online weighted resampling (OWR) technique to perform personalized BP prediction in an online manner. The Bootstrap-based OWR technique is devised to provide a dynamic resampling mechanism of historical samples to conquer possible concept drift and anomaly points by assigning different weights. Specifically, based on the prediction error of the incoming sample (e_t) and all historical samples ($\bar{e}_t$) , OWR employ three types of strategies for tuning sample weights as follows:

(1) Anomaly adaption: if the prediction error of x_t is significantly larger than the mean prediction error of historical samples, its weight will be reduced:

$$w_t = \begin{cases} \alpha & \text{if} \quad e_t > \varepsilon \bar{e}_t \\ 1 & \text{else} \end{cases}, \tag{9}$$

where $\varepsilon > 1, \alpha < 1$.

(2) Concept drift adaption: the weight of samples before the warning period t_w will be reduced when concept drift if confirmed:

$$w_{t':t'<t_W} = \begin{cases} \beta \cdot w_{t'} & \text{if} \quad w_{t_W} > L_W \quad \text{and} \quad w_{t_D} \geq L_D \\ 1 & \text{else} \end{cases}, \tag{10}$$

where $\beta < 1$, t_W, t_D denote the warning time and drift confirmed time, with L_W and L_D the corresponding threshold error values, $t_W < t_D$, $L_W < L_D$.

(3) Forgetting mechanism: the weight of all historical samples is scaled down whenever a new sample x_t arrivals:

$$w_{t'} = \gamma \cdot w_{t'}, t' < t. \tag{11}$$

Through OWR, the samples reflecting current environment/concept will be more likely appear in bootstrap dataset for model update.

3.2 Taxonomy 2: traditional machine learning methods vs. Deep learning methods

According to whether feature extraction and model building are performed jointly, related work can be divided into two folds, namely *traditional featured-based ML methods*, and *DL methods*. Figure 6 presents a generalized pipeline for data-driven BP prediction. In ML methods, tedious feature engineering (including feature construction & extraction, and feature selection/transformation) has to be performed ahead of model training to define and screen out the most informative features that related to prediction task. Therefore, related works are mainly focused on signal processing, multi-sensor fusion, feature exploration, and feature screen, etc. In DL methods, feature engineering is no longer necessary due to the powerful capability of DL in learning complex representations as well as relationships directly from raw data, which enables end-to-end training. Therefore, related work mainly focuses on adapting classical model from other domains such as computer vision, etc., or designing specific models to improve prediction performance.

3.2.1 Machine learning-based methods

The current reviews mainly introduce relevant articles one by one in an exhaustive manner. Herein, according to the BP prediction pipeline illustrated in Fig. 6, we will decouple and summarize relevant work in turn from the following aspects, namely *signal denoising*, *segmentation*, *data cleaning*, *peak detection*, *feature extraction*, *normalization*, *feature selection/reduction*, *training algorithms*, and *hyper-parameter optimization and model selction*.

Signal denoising The signals collected from sensors are usually disturbed by all kinds of noises. Denoising signals is a pre-step for feature point positioning and feature extraction. Specifically, ECG signal is disturbed by power line interference (PLI), baseline wandering (BW), motion artifacts (MA), muscle contractions/artifacts, instrumental and electrosurgical noise (Butt et al. 2015; Joshi et al. 2013). Similarly, PPG signal contains PLI, BW, MA, low amplitude PPG signal, etc (Mishra and Nirala 2020). We group signal denoising

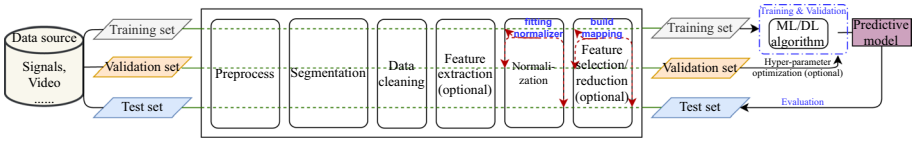

Fig. 6 A general pipeline for blood pressure prediction

methods into four types of *time domain based*, *frequency domain based*, *time-frequency domain based*, and *DL based*. A simple comparison of the popular signal denoising methods is presented in Table 5. Abderahman et al. (2017) proposed a novel method based on EMD to suppress transit MA and MA randomness in Oscillometric waveform signal.

Segmentation The preprocessed raw signals are segmented (refer Sect. 5.1.3) into disjoint segments, each segment corresponds to a sample, which is the basic unit for training and validation. Note that input signal (e.g PPG, ECG, etc.) and the corresponding ABP signal are performed synchronously.

Data cleaning Data cleaning is an indispensable step to improve the quality of data used for training model, especially in intensive care unit (ICU) patient's data, since the database contains signals disturbed by all kinds of noises and even irregular waveform influenced by sensor position movement or change. Currently, the popular methods used for data cleaning are rule-based. Specifically, several metrics are used as indicator to evaluate the signal quality, and the signal segments with value out of the reasonable range of these metrics are identified as invalid signal. Usually used metrics include Skewness (Liang et al. 2018; Qin et al. 2021), BP range limitations (Baek et al. 2019; Xing and Sun 2016; Schrumpf et al. 2021b; Schlesinger et al. 2020; Zhang et al. 2021a; Harfiya et al. 2021; Zhang et al. 2021a), periodicity check (Leitner et al. 2021), sanity checks and consistency check of signal segments (Baker et al. 2021; Baek et al. 2019; Xing and Sun 2016), etc. Besides, a few authors tried to identify invalid signal with the aid of classifier. For example, Monte-Moreno (2011) additionally trained a linear classifier to distinguish the "no signal" (corruption/loss of signal, background noise) from normal signal. This, however, increases the cost of labeling samples.

Peak detection Peak detection is a crucial prerequisite step to accurate physiological feature extraction. Specifically, a standard PPG cycle contains five key points, namely onset, systolic peak, valley, dicrotic peak, and offset. A standard ECG cycle contains five key points, marked as G, Q, R, S, and T, where R peak is the most important and the most recognizable peak. Table 6 summarized several popular peak detection algorithms of signal.

Feature extraction Feature extraction is critically important step of conventional feature-based methods for BP prediction. Note that the concept of *whole-based features* that appeared in several literatures (Kachuee et al. 2016; Mousavi et al. 2019b) means time domain signal in a specific interval. In other words, there is no feature extraction actually.

Since both PPG features and ECG features responsible for BP prediction have been extensively explored and confirmed in several representative literatures (Chowdhury et al. 2020; Monte-Moreno 2011; Miao et al. 2019; Yang et al. 2020a; Kachuee et al. 2016; Thambiraj et al. 2020; El-Hajj and Kyriacou 2021b; Lin et al. 2021a; Ding et al. 2019; Maqsood et al. 2022), we will not detail these features trivially. Summarily, according to Miao et al. (2019), these feature can be grouped into two types of physiological features and informative features. Physiological features are defined based on feature points of raw signals with physiological meanings, while informative features are the representation of the whole signal reflecting some properties of the signal. In addition, demographic features are usually used as supplement to extracted feature to improve the prediction accuracy of the model. A summary of these types of features is presented in Table 7. Besides, the exploration of new features has never stopped, and Table 8 summarized several novel features proposed in related literatures recently.

As mentioned before, the utilization of physiological features for BP prediction has been thoroughly investigated. However, the extraction of physiological features rely on precise positioning of feature points, which may be very difficult in ICU patient's data or high BP patients with diversified even deteriorated PPG morphology and disturbed by all kinds of

Table 5 Summary of methods for denoising signal

Type	Methods	Details	Function	Characteristic	References
Time domain based	Kalman filter Welch et al. (1995)	Estimate the state variables (BW) using a polynomial approximation independent of the signal	Remove BW	Fails for the condition under high frequency changes	Kurylyak et al. (2013)
	Mean filter	Smooth signal using window-based averaging operation	Remove BW	Simple	Paviglianiti et al. (2020a)
	Median filter	Smooth signal using window-based median operation	Remove BW	Simple	Senturk et al. (2020)
	Cubic spline	Smooth signal using cubic spline interpolation	Remove BW	Simple	Chen et al. (2019)
	Savitzky-Golay filter	Smooth signal based on local polynomial least squares fitting	Remove high-frequency noise	Smooth signal while keep the shape of the signal unchanged	Xing and Sun (2016), Esmaelpoor et al. (2021a), El-Hajj and Kyriacou (2021a)
	Hampel filter	Detect outlier based on statistical analysis	Remove outlier	Overcomes the problem that the previous filters are sensitive to outliers	Slapničar et al. (2019), Rong and Li (2021a), Liu et al. (2020a)
Frequency domain based	FIR Das and Chakraborty (2017)	Includes Kaiser window and other window methods such as Hamming window and Hanning window, etc. can be used to design low/high/band-pass filters and notch filter	Remove muscle artifacts using low-pass filter, BW using high-pass filter, PLI using notch filter	Linear time invariant (LTI) filters	Esmaelpoor et al. (2021a), Thambiraj et al. (2020), Thambiraj et al. (2020), Fong et al. (2019), Zhang et al. (2019a)
	IIR Das and Chakraborty (2017)	Includes Butterworth and Chebyshev, etc.	Remove high-frequency noise	Linear time invariant (LTI) filters, recursive filters	Chowdhury et al. (2020)
	FFT-based	Based on FFT and inverse FFT	Remove low-frequency (e.g. BW) and high-frequency noises	Suitable to stationary signals	Mousavi et al. (2019b), Hsu et al. (2020)

Table 5 (continued)

Type	Methods	Details	Function	Characteristic	References
Time-frequency domain based	DWT	Decompose signal using wavelets (e.g symlet, daubechies, etc.) which removed low-frequency and high-frequency components while reconstruction	Remove BW and high frequency noises such as PLI and muscle artifacts, etc.	Suitable to non-stationary signal	Fan et al. (2021), Kachuee et al. (2016), Zhang et al. (2021a)
	EMD Huang et al. (1998), EEMD Wu and Huang (2009)	Decompose signal into a sum of intrinsic mode functions (IMFs) which removed lower and higher IMFs while reconstruction, EMD is a part of HHT Huang et al. (1998)	Remove muscle artifacts (corresponds to lower IMFs) and BW (corresponds to higher IMFs)	Suitable for non-stationary and nonlinear time series	Hui et al. (2020), Sadrawi et al. (2016), Huang et al. (2019)
	VMD Dragomiretskiy and Zosso (2013)	Decompose signal into an ensemble of band-limited intrinsic mode functions (IMFs), then the signals are reconstructed by ignoring the IMFs corresponding to noise part	Remove BW and high-frequency noises	Adaptive, variational method, the best way to handle noisy input signal, better handle different signal-to-noise ratio regimes and nonstationary artifacts, robust to sampling	Sharifi et al. (2019)
Learned filters	DL methods based	A model is trained in supervised mode to output clean signal given input	Remove all kinds of noise	Lack mathematical theory	Singh and Kaur (2013), Arsene et al. (2019)

IIR infinite impulse response, *FIR* finite impulse response, *FFT* fast fourier transform, *DWT* discrete wavelet transform, *EMD* empirical mode decomposition, *EEMD* ensemble EMD, *VMD* variational mode decomposition

noise (Mousavi et al. 2019b; Haddad et al. 2021). Therefore, some authors try to achieve BP prediction using informative features only or other implicit feature extraction techniques such as K-SVD (Aharon et al. 2006), etc. A summary of these studies in presented in Table 9.

Normalization Normalization is a technique to eliminate dimensional differences between different features. Common normalization techniques include Z-Score standardization, min-max scaling, etc. However, there are two new minor changes when it is applied to DL-based BP prediction. First, for DL methods with raw signal as input, normalization is performed at segment level instead of feature level to eliminate the difference of signals among different individuals (Fan et al. 2019; Slapničar et al. 2019; Schrumpf et al. 2021a; Fan et al. 2021). Second, normalization is also performed on target variable (we call *normalize target*) for better gradient-based update in addition to input. This technique is widely used, especially in those studies of trying to reconstruct ABP waveform (Mahmud et al. 2022; Aguirre et al. 2021; Athaya and Choi 2021; Cheng et al. 2021; Qin et al. 2021). Especially note that when *normalize target* is used for model training, the scaled prediction of the model has to be inversely normalized again in the test stage.

Feature selection/reduction To reduce the number of features and filter out effective features required for model training, there are two solutions. The first is feature selection. Specifically, it can be divided into filter methods, wrapper methods, and embedded methods (Yang et al. 2020a; Chandrashekar and Sahin 2014). For filter methods, variable selection is performed using variable ranking techniques. For wrapper methods, the predictor's performance is served as objective function to evaluate variable feature subset. For embedded methods, feature selection is performed during model training. The second is feature reduction, which represents a class of unsupervised methods that based on some transformation such as principal components analysis (PCA), etc. Table 10 summarizes widely used feature selection/reduction techniques in the relevant literatures.

Besides, Miao et al. (2019) proposed a novel spectral analysis-based weakly supervised feature selection (WSFS) method specific to oscillometric method. Hassani and Foruzan (2019) firstly use ANN module to map the hand-crafted feature set to acquire the so-called latent features, before model training. Herein, ANN actually plays the role of feature reduction. In several studies, partial least square (PLS) model is usually employed to eliminate the multicollinearity issues between variables (Fujita et al. 2019; Singla et al. 2020b).

Training algorithms Different from traditional explicit analytical models such as PTT methods and other haemodynamic-based methods (Liu et al. 2020a; Yamakoshi et al. 2021; Thambiraj et al. 2019; Matsumura et al. 2018; Ebrahim et al. 2019; Hassani and Foruzan 2019; Ganti et al. 2021) where the relationship between explanatory variables and BP is pre-determined. In feature-based ML methods, each algorithm actually determines a hypothesis space, and the algorithm is employed to learn the best mapping between variables and BP from the hypothesis space using training data. Therefore, ML methods enables the model with stronger nonlinear expression ability.

Currently, almost all kinds of classical ML algorithms have been employed to train prediction models. The usage of different algorithms is summarized in Table 11. LR is widely used in this area owning to its strong interpretability and ease-of-use. It is noted that PTT methods (Mukkamala et al. 2015; Peter et al. 2014; Sola et al. 2013) and its variants (Chandrasekhar et al. 2020; Esmaili et al. 2017; Shao et al. 2020; Hsieh et al. 2016; Das et al. 2020) can be seen as a special application of LR algorithm where PTT-related features are served as explanatory variable and linear/quasi linear relationship between these features and BP is assumed. Since different algorithms have different hypothesis spaces, and it is impossible to know which algorithm will derive the best predictor given specific datatset.

Table 6 Several popular peak detection algorithms

Method	Description	Characteristic	Package	Usage reference
AMPD Scholkmann et al. (2012)	Identify peaks by calculating the local maximum scale value	No free parameter values need to be chosen by user, high efficiency, suitable noisy periodic and quasi-periodic signals	Python: ampdLib[a], pyampd[b]	Detect QRS peak in ECG Scholkmann et al. (2012), detect systolic peak in PPG (Mousavi et al. (2020)
DPI Ramakrishnan et al. (2014)	DPI is applied on the high-pass filtered signal to detect R-peak	Does not make use of any threshold or differencing operation	C++, Matlab: QRS-detection[c]	Detect QRS in ECG signal Esmaelpoor et al. (2020)
Wavelet transform Li et al. (1995)	Based on quadratic spline wavelet	Can identify the five points-P-Q-R-S-T in ECG signal, robust to BW and noises, suitable for processing time-varying biomedical signal	–	Detect QRS in ECG Li et al. (1995), detect main peak in PPG Lin et al. (2020)
CWT-based Du et al. (2006)	Identify peaks by applying CWT -based pattern matching	Can identify peaks across scales and with varying intensities in a single spectrum	Python: scipy.signal.find_peaks_ cwt	Detect R-peak in ECG signal
find_peaks	Detect peaks based on peak properties	Many parameters need to set manually	Python: scipy.signal.find_peaks , Matlab: findpeaks	Detect systolic peak in PPG Thambiraj et al. (2020)
DL methods Vijayarangan et al. (2020)	A network is trained to detect peaks	High accuracy, needs a lot of labeled data for training	RPNet[d]	–

AMPD automatic multi-scale-based peak detection, *DPI* dynamic plosion index, *CWT* continuous wavelet transform

[a]ampdLib: https://github.com/LucaCerina/ampdLib

[b]pyampd: https://github.com/ig248/pyampd

[c]QRS-detection: https://github.com/AgnieszkaZadlo/QRS-detection

[d]RPNet: https://github.com/acrarshin/RPNet

Therefore, a common practice (Yang et al. 2020a; Miao et al. 2017; Kachuee et al. 2016; Monte-Moreno 2011; Esmaelpoor et al. 2021a; Hasanzadeh et al. 2019; Thambiraj et al. 2020; Chowdhury et al. 2020; Chen et al. 2019) is that a prediction model is trained using each algorithm on the given dataset, respectively, and then the prediction model with the best performance is selected as the final model for further test. It is observed that SVR, AdaBoost and RF has gained the most popularity due to their excellent performance in most cases (Zhang et al. 2021a).

Hyper-Parameter optimization and model selection During the construction of prediction model, there are two type of parameters that need to be determined, namely model parameters and hyper-parameters. For example, SVM algorithm contains model parameters (i.e weights and bias) and hyper-parameters such as penalty coefficient C and coefficient of kernel function γ, etc. Model parameters are tuned iteratively based on training set during training. Hyper-parameters are determined using hyper-parameter optimization (HPO) techniques, based on validation set. Currently used HPO techniques include grid search, random search, genetic algorithm (GA), particle swarm optimization (PSO), Bayesian optimization (BO), etc. All of them are black-box optimization methods, and a summary of these HPO methods is presented in Table 12.

3.2.2 Deep learning-based methods

Deep learning based methods for BP prediction are mainly characterized in neural network architecture exploration. Generally speaking, there are two directions are available for consideration, (1) the first direction is to utilize classic models from other fields such as computer visions (CV), etc. However, CV model is usually designed based on 2D-convolution, while physiological signal is one-dimensional. In practice, there are two ways to overcome this problem. The first way is to convert 1D signal in some way to a 2D format, and then CV models can be used directly. The second way is to modify 2D-convolution-based CV models to its corresponding 1D-convolution-based format. (2) the second direction is to design domain-specific models. Next, we will review related work based on the model architecture used. Relevant studies can be coarsely divided into two parts, i.e *basic DL models for BP prediction*, and *hybrid DL models for BP prediction*.

3.2.3 Part 1-Basic DL models for BP prediction

As shown in Fig. 7, we summarized five basic model structures for BP prediction, namely *FFNN*, *RNNs*, *CNN*, *BNN*, and *Siamese architectrue*, of which the first three are widely used and the latter two are rarely used.

FFNN Feed-forward neural network (FFNN), also known as artificial neural network (ANN), multilayer perceptron (MLP) or back-propagation network, is one of the easiest-to-understand neural networks that has been widely used, especially in earlier studies (Jeong et al. 2019; Attarpour et al. 2019; Maher et al. 2021; Yin et al. 2021; Xing and Sun 2016; Wang et al. 2018a; Zhang and Wang 2017; Wang and Zhang 2017; Sadrawi et al. 2016; Wang et al. 2018b; Tan et al. 2018; Wu et al. 2016; Mahmud et al. 2022; Lin et al. 2021b). A classical FFNN usually contains three layers of input layer, hidden layer and output layer, and any two neuron nodes of adjacent layers are connected with each other. For example, Wang et al. (2021) built a stacked autoencoder for the estimation of BP under blood loss, the network is firstly perform unsupervised layer-by-layer greedy pre-training, and then is

Table 7 Summary of different types of features

Type	Detail	Description
Physiological features	Pulse transit time	Time duration between R-peak of ECG and different feature points of synchronous PPG signal
	Time duration	Time difference between different feature points of signal
	Amplitude feature	Amplitude of/between different feature points of signal
	Pulse width	Pulse width at different level of signal amplitude
	Area	Area between different feature points under signal curve
	Combined features	Defined based on ratio and difference operations, such as K-value Miao et al. (2019), etc.
Informative features	Statistical features	Mean, standard deviation, Skewness Liang et al. (2018), Kurtosis Miao et al. (2019), signal mobility and signal complexity Simjanoska et al. (2018, 2020), etc.
	Frequency domain based features	Amplitude of specific frequency component of signals, spectral entropy Monte-Moreno (2011)
	Wavelet domain based features	Qualify the complexity of signals, include wavelet energy entropy Miao et al. (2019)
	Entropy based features	Describe the confusion degree of signals, include sample entrope and approximate entropy Miao et al. (2019)
Whole-based features	–	Signal segment in a specific interval
Demographical features	–	Age, sex, height, BMI Li et al. (2017), etc.

further trained as a whole through supervised learning. Zhang and Wang (2017) built a BP network for BP estimation, where genetic algorithm is firstly used to initialize model parameters, and then the model is trained by back-propagation.

Recently, Huang et al. (2022) develops a BP prediction model based on the novel MLP-Mixer architecture. As Fig. 8 illustrates, the multi-filter to multi-channel (MFMC) technique is firstly used to extend the channel dimension of the raw input signal, and then the embedded representation is fed into the MLP-Mixer module which iteratively performs channel-mixing and temporal-mixing based solely on MLP module. Inspired by biological neurons, Ji et al. (2022) proposed a novel dendritic neural model (DNM), a single neural model with a plastic dendritic structure, for BP prediction. Experiments show that this

Table 8 Several novel features proposed in related studies

Name	Definition	Description
PI Shin and Min (2017)	$\frac{l_{NN}+l_{NP}}{l_{PN}+l_{NN}+l_{NP}}h$	Arrival time or velocity related parameter of the reflected wave, based on the polarities of PPG's derivatives
Womersley number Thambiraj et al. (2019, 2020)	$\alpha = \sqrt{\frac{w\rho}{h}}$	The ratio of inertial force (pulsatile frequency) to viscous effects
PIR Ding and Zhang (2015), Ding et al. (2017)	$\frac{1}{\alpha}\ln\frac{I_H}{I_L}$	The ratio of PPG peak intensity I_H to valley intensity I_L
LCFs Fujita et al. (2019)	LCF 1, LCF 2	Number of crossings between standardized APG waveform and a contour line, total length of each contour line lying within the waveform
Fast upstroke time intervals Natarajan et al. (2021)	Finger b-time, ear "STT"	PPG foot to minimum second derivative time interval, PPG amplitude divided by maximum derivative

PI pressure index, *PIR* photoplethydmogram intensity ratio, *LCFs* level-crossing features

Table 9 Implicit feature extraction approaches appeared in related studies

Name	Description
FFT based Xing and Sun (2016)	Amplitude and phase features are extracted from PPG cycle based on FFT transform
DCT based Wang and Zhang (2017)	The first 15 points of the DCT transform sequence of raw PPG signal is used as input feature
DWT based Gao et al. (2016)	DWT coefficients of raw PPG segments is used as input feature
MSE Sadrawi et al. (2016)	75 scale of MSE of raw PPG segments is used as input feature
SCSA Li and Laleg-Kirati (2021)	SCSA features are derived by decomposing PPG segments into two partial sums
K-SVD based Bose and Kandaswamy (2018)	Feature extraction is modeled as a dictionary learning problem, the sparse features of PPG cycle generated based on K-SVD is used as input feature
McSharry dynamic model fitting based Mousavi et al. (2019a)	Feature extraction is modeled as a signal fitting problem, the parameter values in McSharry equations by fitting ECG signal based on McSharry Dynamic model is used as input feature
Autoencoder based Shimazaki et al. (2018), Mahmud et al. (2022)	The outputs of autoencoder which is trained by reconstructing input signal or converting input signal to target signal, are used as input feature

FFT fast Fourier transform, *DCT* discrete cosine transform, *DWT* discrete wavelet transform, *MSE* multiscale entropy, *SCSA* semi-classical signal analysis

special model has achieved competitive results both on static BP estimation and long-term BP estimation.

RNNs Considering the temporal order and dynamic nature of physiological signals, standard recurrent neural networks (RNNs) and its variants such as GRU (Chung et al. 2014), LSTM (Hochreiter and Schmidhuber 1997), and Bi-LSTM (Liwicki et al. 2007), etc., have been widely used for BP prediction due to their strong ability to model temporal dependencies (El-Hajj and Kyriacou 2020a; Liu et al. 2018; Koshimizu et al. 2020; Li et al. 2017, 2020a; Lee et al. 2021; Paviglianiti et al. 2020b; Lo et al. 2017). Li et al. (2017) proposed a LSTM model with contextual layer (named LSTM-CL) to better predict individual's BP using both clinical data and contextual data. Su et al. (2018) proposed a deep LSTM model for long term BP prediction. Similar to Su et al. (2018), Li et al. (2020a) proposed a deep LSTM model with residual connection for BP prediction, and investigated the best model configuration in terms of both network depth and residual connection. Results indicate that the average prediction accuracy decreases with the increase of network depth, which is unexpectedly opposite with the conclusion in study (Su et al. 2018). Furtherly, El-Hajj and Kyriacou (2021a, b) proposed a deep LSTM-based network with attention mechanism to predict BP. Specifically, the attention module attached is employed to focus on the more important hidden states in each time step automatically.

CNN Generally, 1D convolution network is utilized to learn temporal features/patterns from raw signal (Esmaelpoor et al. 2021b; Baek et al. 2019, 2020; Athaya and Choi 2021; Sadrawi et al. 2020; Cheng et al. 2021; Ibtehaz and Rahman 2020). For example, Baek et al. (2019) built a fully-convolution network based on the proposed extraction-concentration block (EC_block) for BP prediction. As Fig. 9 illustrates, the network comprising

Table 10 Several feature selection/reduction approaches appeared in relate studies

Type		Name	References
Feature selection	Filter method	PCC	Ding et al. (2019), El Hajj and Kyriacou (2020a), Rong and Li (2021b)
		MI	Ding et al. (2019)
		MIC	El-Hajj and Kyriacou (2021b)
		MIV	Chen et al. (2019), Tan et al. (2018), Zhang and Wang (2017)
		Relief	Chowdhury et al. (2020), Slapničar et al. (2018)
		GA	Miao et al. (2017)
	Wrapper method	RFE	El-Hajj and Kyriacou (2020a, 2021b)
		Sequential forward selection	El Hajj and Kyriacou (2020a)
		Moving background algorithm	Attarpour et al. (2019)
		Binary GA	Thambiraj et al. (2020), Attarpour et al. (2019), Dagamseh et al. (2021)
	Embedded method	RF	El Hajj and Kyriacou (2020a)
		RFFS	Chiang and Dey (2018), Chiang et al. (2021)
	Hybrid	PCC + MIC + RFE	El-Hajj and Kyriacou (2021b)
		PCC + Sequential forward selection + Random forest feature importance + RFE	El-Hajj and Kyriacou (2021a)
		Fast two-stage feature selection method	Lin et al. (2021a)
Feature reduction	–	PCA	Dagamseh et al. (2021), Mousavi et al. (2019b), Yang et al. (2020a), Kachuee et al. (2016), Luo et al. (2019)
Hybrid	–	Model-based feature selection + Covariance- maximizing dimensionality reduction	Yousefian et al. (2020)

PCC Pearson correlation coefficient, *MI* mutual information, *MIV* mean impact value, *MIC* maximum information coefficient, *RFE* recursive feature elimination

Table 11 Representative machine learning algorithms used for BP prediction

Algorithm	References
LR	Natarajan et al. (2021), Haddad et al. (2021), Dey et al. (2018), Esmaelpoor et al. (2021a) Chowdhury et al. (2020), Yousefian et al. (2020), Zhang et al. (2021a),Zheng and Yu (2021), Lazazzera et al. (2019), Singla et al. (2020a), Khalid et al. (2018),Marzorati et al. (2020)
PLS	Zhang et al. (2021b), Fujita et al. (2019), Singla et al. (202b0)
GPR	Zheng and Yu (2021), Huttunen et al. (2019), Chowdhury et al. (2020), Chen et al. (2022),Esmaelpoor et al. (2021a)
SVR	Fong et al. (2019), Khalid et al. (2018), Zheng and Yu (2021), Esmaelpoor et al. (2021a),Chowdhury et al. (2020), Hassani and Foruzan (2019), Kachuee et al. (2015), Chen et al. (2019),Zhang et al. (2017, 2019b, 2021a), Dagamseh et al. (2021)
AdaBoost	Hasanzadeh et al. (2019), Ibrahim and Jafari (2019), Kachuee et al. (2016), Wang and Zhang (2017),Mousavi et al. (2019b, 2020), Zhang et al. (2021a)
RF	Kachuee et al. (2016), Simjanoska et al. (2018, 2020), Chen et al. (2021), Fati et al. (2021) Huang et al. (2019), Thambiraj et al. (2020), Xing et al. (2019), Monte-Moreno (2011),Bose and Kandaswamy (2018), Zhang et al. (2021a), He et al. (2016a)
CART	Golino et al. (2014), Zhang et al. (2018), Chiang Chiang et al. (2021), Esmaelpoor et al. (2021a)
DT	Slapničar et al. (2018), Khalid et al. (2018), Chowdhury et al. (2020)

two branches for learning time domain and frequency domain information of raw signal, respectively. The idea behind EC_block is that multiple dilated convolution are performed to learn various relationships between different neighboring pixels and the concatenated output is reduced to its initial dimension through 1×1 convolution, then strided convolution is performed to increase the depth of the features and halve the temporal length. Besides, a few authors (Sasso et al. 2020; Wang et al. 2020) directly use classical 2D convolution-based CV networks for BP estimation by converting PPG segment to 2D format in some way. Qiu et al. (2021) built a composite 2D-CNN based model for BP prediction, where raw signals (PPG &ECG) are firstly processed through CNN-Sequential-Adapt layer to generate appropriate output for ResNet-25 with channel attention. Malayeri and Khodabakhshi (2022) built a two stream compound CNN model for BP prediction using PPG signal. As Fig. 10 illustrates, the 1D CNN module learns morphological information directly from raw PPG segment, and the 2D CNN module learns information from the 2D image converted from the raw PPG segment based on fuzzy recurrent plot (FRP), then the information from the two streams are fused for BP prediction. In addition, considering the duration of sample and network complexity, sampling techniques (Qiu et al. 2021; Baek et al. 2019; Panwar et al. 2020) are usually used to down sample the raw signals before it is fed into the network.

BNN Boosting neural network (BNN) (Schwenk and Bengio 2000) is the product of the combination of the general Boosting algorithm and neural networks. As Fig. 7d illustrates, BNN is a class of networks with cascade structure, and the usually used decision tree serving as base learner in Boosting algorithm is replaced with neural network. Song et al. (2019, 2021) proposed a stacked DNN model for BP prediction using both PPG &ECG signals and demographical features. As Fig. 11 illustrates, a shallow FFNN module is trained using the total features for BP prediction in the first stage, and then a second FFNN is trained using the total features and the estimated BP value of the last model as input. In other words, the second model is trained by fitting the residual between the last

Table 12 Summary of several hyper-parameter optimization techniques

Type	Method	Description	Characteristic	Package	References
Sampling based	Grid search	Elaborately evaluate the Cartesian product of the values of the hyper-parameters set	The basic method, curse of dimensionality	Python: scikit-learn	Zhang et al. 2018, (2019b), Zheng and Yu (2021), Simjanoska et al. (2020), Gao et al. (2016)
	Random search Bergstra and Bengio (2012)	Randomly select the next candidate for evaluation until stop condition is triggered	Alternative to *grid search*	Python: scikit-learn	Ji et al. (2022)
Population based	Genetic algorithm Whitley (1994)	Maintains a population, and improve the population using several sequential operations (selection, crossover and mutation) inspired from biological evolution	Suitable for both discrete and continuous problems	Python: Geatpy[a], scikit-opt[b]	(Sadrawi et al. (2020), Tan et al. (2018), Chen et al. (2019, 2021)
	Particle swarm optimization Kennedy and Eberhart (1995)	Similar to *genetic algorithm*	Suitable for continuous problem	Python: scikit-opt[b]	Fan et al. (2019)
Others	Bayesian optimization Snoek et al. (2012)	An acquisition function is used to decide which point to evaluate next by trading off exploration and exploitation	Global optimization method	Python: bayes_opt	Slapničar et al. (2018), Franco et al. (2019)

[a] http://geatpy.com/

[b] https://github.com/guofei9987/scikit-opt

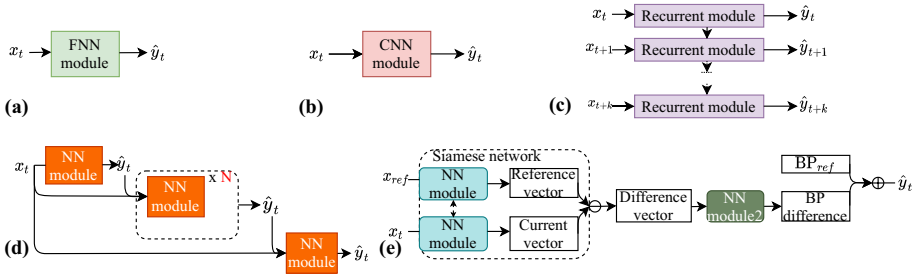

Fig. 7 Several representative network architectures for blood pressure prediction. **a** feedforward neural networks (FFNN); **b** convolutional neural networks (CNN); **c** recurrent neural networks (RNN); **d** boosting neural networks (BNN); **e** siamese architecture

model's output and the genuine BP value. The so-called stacked DNN is actually a BNN, although it is not explicitly stated in the paper.

Siamese architecture Siamese network (Bromley et al. 1993; Chopra et al. 2005) is a novel architecture that is closely related to contrastive learning and representation learning, and has been widely used in various computer vision tasks. Generally, Siamese is composed of two identical networks and one cost module. The network accepts a pair of samples as input, and the outputs of the two sub-networks are passed to the cost module to compute similarity. Schlesinger et al. (2020a, b) firstly proposed a novel Siamese-based model for BP prediction. Specifically, as Fig. 7e illustrates, they made two modifications, (i) it is a regression network: the final layer of the network is fully-connected layer with linear activation, which outputs BP difference; (ii) instead of metrics with positives value used in cost module, the two resulting feature vectors are directly subtracted. In other words, the model learn the difference vector of current input with respect to reference input, which is utilized to estimate BP difference. The final BP is acquired by plus the estimated BP difference with the reference BP.

3.2.4 Part 2-Hybrid DL models for BP prediction

The combination of multiple basic model architectures is also widely used in related studies for BP prediction. As shown in Fig. 15, we summarized a total of three combination modes of designing hybrid architectures that are widely used in related studies.

In *mode-1* (Fig. 15a), CNN module is followed by RNN module, which is just opposite to mode-1. Baker et al. (2021) built a hybrid model where temporal CNN module is firstly used to identify important features and patterns, which is followed by Bi-LSTM module for modeling temporal dependency. Panwar et al. (2020) use the similar network as in Baker et al. (2021) for estimating BP and heart rate simultaneously. Esmaelpoor et al.

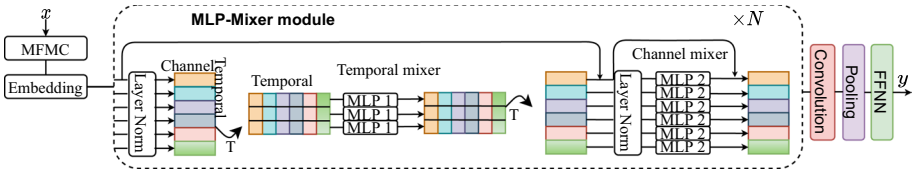

Fig. 8 MLP-Mixer based model for blood pressure prediction proposed in study Huang et al. (2022)

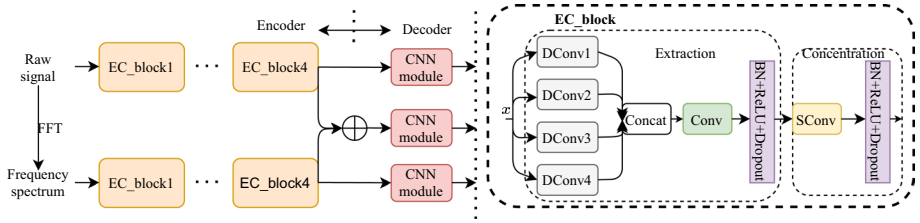

Fig. 9 Fully-convolution network for blood pressure prediction proposed in study Baek et al. (2019). *DConv* dilated convolution, *SConv* strided convolution

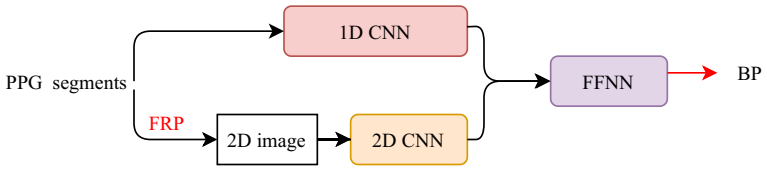

Fig. 10 Two stream compound CNN network for blood pressure prediction proposed in study Malayeri and Khodabakhshi (2022). *FRP* fuzzy recurrent plot

(2020) proposes a novel two-stage hybrid model for BP prediction where both temporal dependency in each task and correlations between SBP and DBP prediction are modeled. As Fig. 12 illustrates, in the first stage, two separate sequential CNN modules are trained for SBP and DBP prediction, respectively. In the second stage, for each prediction task, in addition to the resulting feature vector of the first stage, both the estimated BP in the first and the second stages from another task is used to train a LSTM model. Leitner et al. (2021) proposed a transfer learning framework for BP prediction based on the hybrid CRNN model. As Fig. 13 illustrates, a pretrained/generalized model is firstly trained using source patient's data and is then fine-tuned using partial target patient's data to obtain personalized model. Experimental results indicate that the best performance is obtained when retraining only the last fully-connected layer and the last convolutional layer. Eom et al. (2020) proposed a CNN-RNN model with attention mechanism to predict BP. Specifically, a VGGNet-like CNN module is modified to automatically extract features from raw signals, which is followed by a Bi-GRU module to encode temporal information between the learned features. Finally, the estimated BP is computed based the attention module to focus on the different importance of hidden states in each time step. Chuang et al. (2021) proposed a hybrid model similar to Eom et al. (2020) except that the CNN module extracts features from both temporal domain and frequency domain of the input signal.

Different from the above-mentioned literatures (Baker et al. 2021; Panwar et al. 2020; Esmaelpoor et al. 2020; Leitner et al. 2021) where RNN module is used to model temporal dependency within sample (i.e. the output of convolution module), in Jeong and Lim (2021) work, RNN module is utilized to model temporal dependency between samples since the CNN module is time-distributed.

In *mode-2* (Fig. 15b), CNN and RNN modules are used to capture different feature in parallel, and then these features are further fused through FFNN module. For example, Miao et al. (2020) proposed a hybrid model consisting of ResNet-50 and LSTM models to capture morphological and rhythmic features that relates to BP variation, respectively, which is followed by several fully-connected layers for fusion of the two types of features.

Rong and Li (2021a) proposed a hybrid model composed of two CNN modules and one Bi-LSTM module to extract morphological, frequency spectrum and temporal features of PPG signal, respectively. Unlike the above-mentioned studies (Miao et al. 2020; Rong and Li 2021a) that only consider learning from time domain signals, as Fig. 14 illustrates, Slapničar et al. (2019) proposed a hybrid network that learning both in time domain and time-frequency domain of PPG signal and its derivatives for BP estimation.

In *mode-3* (Fig. 15c), time-distributed ANN is followed by RNN module. Tanveer and Hasan (2019) proposed an ANN-LSTM model for BP prediction where a time-distributed ANN module is used to extract morphological features from multiple segments of raw PPG and ECG signals in parallel, which is followed by a LSTM module to learn the time domain variation of the extracted features.

In addition, Yang et al. (2020a) proposed a novel framework for BP prediction based on the hypothesis that estimation performance may be improved by separating BP and feature variations into low frequency components (LFC) and high frequency components (HFC), and modeling each separately. As Fig. 16 illustrates, both BP sequence and feature sequences are separated into LFC and HFC using a first-order low-pass Butterworth digital filter, and then a regression model is trained using feature-BP pair in each frequency component. In the test phase, the sum of the two model's outputs plus the mean BP of test subject constitutes the final BP. Note that although the model uses a multi-branch structure, while with purpose different from previous studies (Miao et al. 2020; Rong and Li 2021a).

Remark There is no one-size-fits-all model that work well on all problems and datasets. For example, Paviglianiti et al. (2020b) finds that the selection of model architecture is dependent on the type of input signal. Xiang et al. (2021) finds that LSTM works best when there is a high temporal dependency between trajectory data, otherwise RF model works best. Currently, model designing and configuration follows certain principles (qualitative) and relies on extensive trial and error. Generally, for models of sequential structure, CNN or FFNN module is usually served as feature learner to extract and fuse features, RNNs module is usually utilized to model temporal dependencies. FFNN is usually used with hand-crafted features (Wu et al. 2016; Tan et al. 2018; Zhang and Wang 2017; Attarpour et al. 2019; Maher et al. 2021; Yin et al. 2021; Wang et al. 2018a, b) or implicit features derived from some transformations (Xing and Sun 2016; Wang and Zhang 2017) or sparse representation (Bose and Kandaswamy 2018) of raw signal as input. Since a single RNNs module has no capability of feature extraction, it is usually used in combination with other modules such as CNN or FFNN, etc. (e.g., Baker et al. 2021; Panwar et al. 2020; Esmaelpoor et al. 2020; Leitner et al. 2021), or directly feed with extracted features (e.g., El-Hajj and Kyriacou 2021b). For models composed of multiple parallel structures, features learned from different streams of network are further fused based on FFNN module for BP prediction. The features to be fused may come from different signal sources (e.g., Baek et al. 2020), different modalities of specific signal (e.g., Rong and Li 2021a;

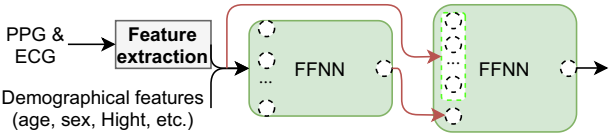

Fig. 11 Boosting neural network for blood pressure prediction proposed in study Song et al. (2019)

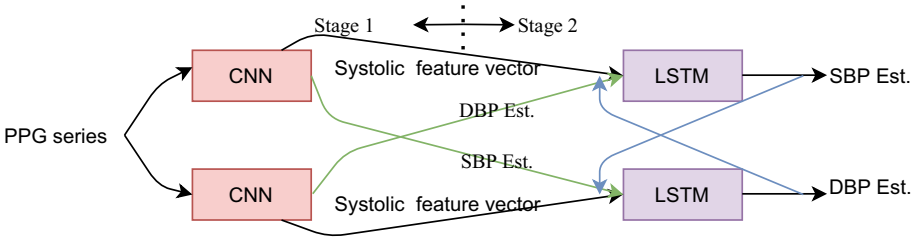

Fig. 12 Model architecture proposed in study Esmaelpoor et al. (2020)

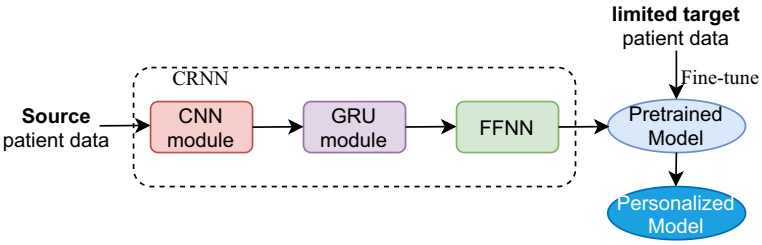

Fig. 13 Transfer learning framework proposed in study Leitner et al. (2021)

Slapničar et al. 2019), or even outputs of different network with the same input (e.g., Miao et al. 2020).

3.2.5 Discussion

In this subsection, we make a general comparison of traditional feature-based ML methods and DL methods for BP estimation from the following four aspects.

Performance Due to the powerful ability of deep learning to extract, represent, and fit complex relationships, it is reasonable to believe that given a sufficiently large dataset, a well-designed, sufficiently complex, and optimized deep neural network model can achieve superior performance over feature-based ML methods. Several studies (Slapničar et al. 2019; Fan et al. 2019, 2021) with empirical evaluation have confirmed this.

End-to-end training property As Fig. 6 illustrates, for feature-based ML methods, feature engineering is an extremely important pre-step for training ML model. However, for DL methods, a predictive model can be obtained by training directly with raw data (only simple preprocess in required) directly in an end-to-end fashion. In other words, automatic feature extraction/learning and model training are performed simultaneously by minimizing the prediction loss at each iteration.

Scalability Generally, for traditional feature-based ML methods, a model is trained using the whole dataset in each iteration, and the model has to be trained from scratch if new data arrives. However, for DL methods, due to the modularity architecture of neural network models with fully parameterized characteristics, DL model is scalable to data size and can be easily used for incremental update. The practice of using pre-trained models for transfer learning actually takes advantage of this characteristic of DL model. Besides, the modularity characteristics of neural network enables the popularity of deep multi-task learning for BP estimation (refer Sect. 3.3.2).

Interpretability Lack of interpretability is a common issue in ML, especially in DL. For traditional feature-based ML methods, hand-crafted features are usually designed with inspiration from physiological explanation or by capturing some characteristics of input data. Although DL method seems to have higher performance than traditional feature-based ML methods, it remains unclear whether DL models learn effective and universal features, and what is the difference and relationship between these features and hand-crafted features. Qin et al. (2021) has visualized the feature representation of the trained convolution-based Autoencoder model for converting PPG signal to ABP waveform, and finds that the encoder learns a sparse, hierarchical abstract of signal segments.

3.3 Taxonomy 3: single-task learning vs. multi-task learning

BP prediction is naturally a multi-task learning question where three prediction tasks of SBP, DBP and MBP share the same input. Figure 17 visually shows the difference

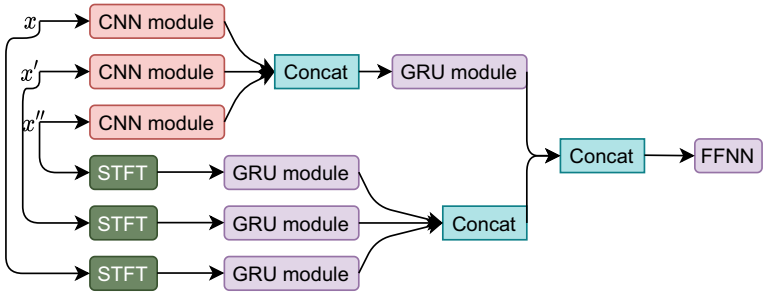

Fig. 14 Model architecture proposed in study (Slapničar et al. 2019)

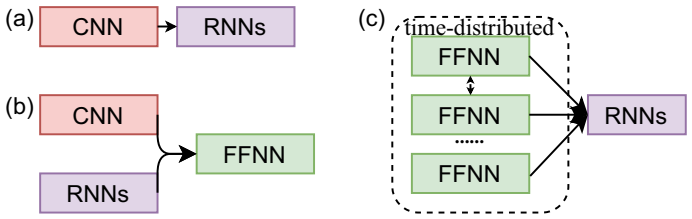

Fig. 15 Several representative hybrid network architectures for blood pressure prediction. **a** mode 1; **b** mode 2; **c** mode 3

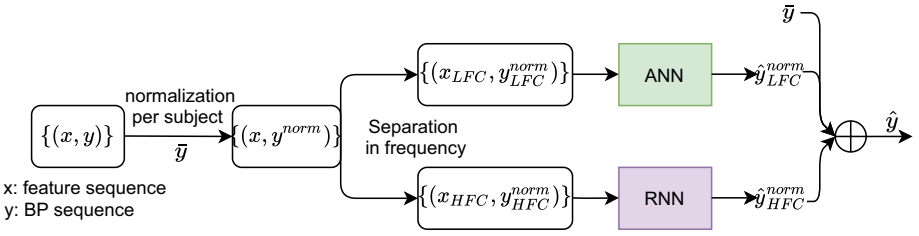

Fig. 16 Model framework for blood pressure prediction proposed in study Yang et al. (2020a)

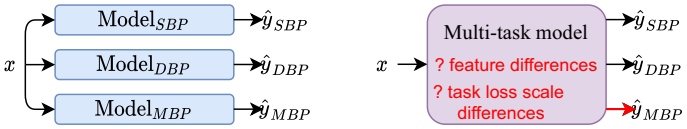

Fig. 17 Single task learning vs. multi-task learning for blood pressure prediction. Left: single-task learning; Right: multi-task learning

between single-task learning (STL) and multi-task learning (MTL) for BP prediction. In STL, a prediction model has to be trained independently for each task. In MTL, while, only one model with multiple outputs is need to be trained for estimating SBP, DBP and MBP in parallel. However, nothing is available for free. Specifically, there are two issues need to be solved in MTL for BP prediction, which will be detailed in the following.

3.3.1 Single-task learning

Almost all of traditional ML-based methods follow the STL mode. Specifically, a model is trained independently for each prediction task, feature selection and model building need to be performed on a task-by-task basis, which is cumbersome and increases memory costs in practical application. Besides, there are several interesting finds as follows by reviewing ML-based methods with explicit feature extraction (El-Hajj and Kyriacou 2021b; Liu et al. 2021; Miao et al. 2017, 2019; Yang et al. 2020a; Ibrahim and Jafari 2019; Song et al. 2019; Yang et al. 2021; Chen et al. 2019). First, there is a large amount of intersection between the feature sets selected for different tasks, and a few features are task-specific. Second, the importance of each feature is varied for different prediction tasks. Third, the prediction error of SBP is significantly larger than DBP. These findings propose new challenges to the designing of accurate MTL model for BP prediction.

3.3.2 Multi-task learning

Currently, studies with respect to multitask learning (MTL) for BP prediction (Tanveer and Hasan 2019; Baek et al. 2019; Eom et al. 2020; Su et al. 2018; Slapničar et al. 2019; Fan et al. 2021, 2019; Esmaelpoor et al. 2020; Zhang et al. 2020b) are all in neuro models owning to the modularity architecture of neural network. In the context of Ruder (2017), almost all of the mentioned studies above except study (Esmaelpoor et al. 2020) follow the hard parameter sharing mode, i.e. several layers are shared among all tasks for learning informative representations, which is followed by multiple independent task networks, with each corresponding to a prediction task. A simplest case is that the number of neurons in the last fully-connected layer equals the number of prediction tasks. What's embarrassing is that a large number of studies such as Slapničar et al. (2019) utilize this simplest MTL network for BP prediction. Surprisingly, there are still a substantial part of DL-based studies (Yang et al. 2021; Baker et al. 2021; Attarpour et al. 2019; Jeong et al. 2019; Wu et al. 2016) where the STL mode is utilized.

Zhang et al. (2020b) proposed a multitask network with adversarial training to predict BP. As Fig. 18 illustrates, additional domain classifier module is introduced to train the feature learner module adversarial to make the learned intermediate features informative (for

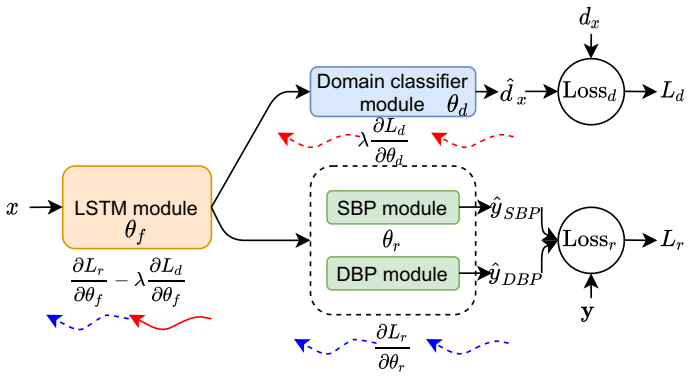

Fig. 18 MTL framework proposed in study Zhang et al. (2020b) for blood pressure prediction

the BP prediction task) as well as cross-individual, allowing for faster knowledge transfer for personalized model. Formally, the feature learner parameterized with θ_f is updated based on gradient as follows,

$$\theta_f \leftarrow \theta_f - \alpha\left(\frac{\partial L_r}{\partial \theta_f} - \lambda\frac{\partial L_d}{\partial \theta_f}\right), \tag{12}$$

where L_r and L_d denote the losses of BP prediction and domain classifier, respectively. The minus sign with red color means feature learner is updated by minimizing prediction loss and maximizing classification loss, trade-off by parameter λ.

However, in comparison with traditional ML-based STL for BP prediction, there are two import issues that we think is critical for achieving successful BP prediction with MTL, under the hard parameter sharing mode. The first question is that *how to conquer the significant loss scale difference among different prediction tasks that may hinder multi-task joint training?* the second question is that *how to cope with the difference in the most informative features accounting for different prediction tasks?*

For the first problem, the multi-task loss can be expressed as $\sum_i w_i \cdot L_i, i \in \{s, d, m\}$, where L_i denotes the loss of task i. The basic idea is to adjust the loss weight $\boldsymbol{w}$ adaptively during training to balance the contribution of each task to the total loss. Up to date, we have found that there are only two articles (Fan et al. 2019, 2021) where there is explicit mechanism to deal with this problem. Fan et al. (2019) proposed a MTL network for BP prediction using ECG signal. As Fig. 19 illustrates, the training process includes two stages: in first pre-training stage, the model is updated based on naive loss-weighting. In the second phase, both model parameters and loss weights are updated alternatively, loss weights is updated heuristically based on PSO optimization. However, the model is complex and the amount of computation is large. Therefore, Fan et al. (2021) further proposed a lightweight version where an adaptive weight learning-based method via the estimation loss trend on validation set is proposed. Specifically, based on the assumption that the trend of task loss variation is positively correlated with the optimization space of the task, the weight of each task loss is defined to be the product of the trend, mean value, and standard deviation of the loss as follows,

$$w_i = T_i^{mean} \cdot T_i^{std} * (L_i^{mean} + L_i^{std}), \tag{13}$$

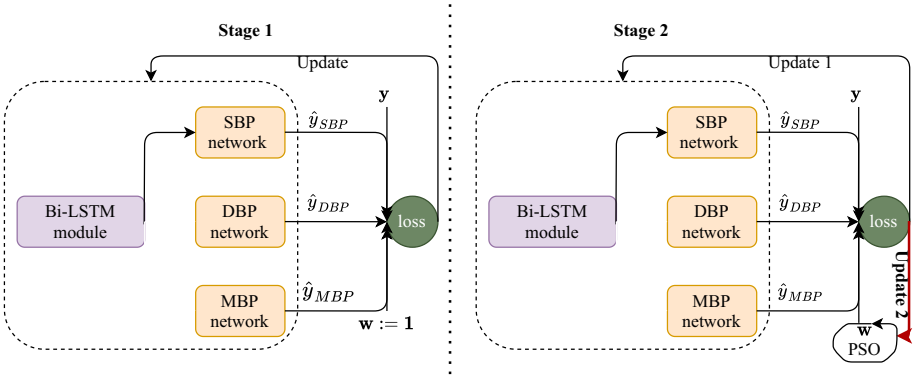

Fig. 19 MTL framework proposed in study Fan et al. (2019) for blood pressure prediction

where L_i^{mean} and L_i^{std} denote the mean value and the standard value of losses on all batches, respectively. T_i^{mean} and T_i^{std} denote the trend of mean value and standard value of losses, which are defined as,

$$\begin{cases} T_i^{mean} = \frac{||L_i^{mean}(k) - L_i^{mean}(k-1)||_1}{L_i^{mean}(k)}, \\ T_i^{std} = \frac{||L_i^{std}(k) - L_i^{std}(k-1)||_1}{L_i^{std}(k)}, \end{cases} \tag{14}$$

where $L_i(k)$ denotes the corresponding statistic value of the k-th epoch. Finally, $\boldsymbol{w}$ is normalized as $w_i = w_i / \sum_i w_i$.

For the second question, almost no relevant research work has been seen. Fan et al. (2019) argue that each BP estimation task has its own characteristics, and developed an attention-based multi-task Bi-LSTM network to automatically select temporal information for SBP, DBP and MBP estimation, respectively. However, no any ablation experiments are performed to validate the effectiveness of the so-called "automatic information selection".

3.4 Taxonomy 4: signal source

Signal source used for BP prediction includes *physiological signal*, *behavior data*, *trajectory data*, and *facial video*, etc.

3.4.1 Physiological signal based

Most studies with respect to BP prediction are based on physiological signals. As Table 13 presents, signals commonly used include *PPG signal*, *ECG signal*, *piezoelectric signal*, *oscillometric wave signals*, and *auscultatory waveform*, etc.

PPG signal-based Unlike that ECG signals are measured at the wrist chest with multiple electrode attached, PPG signal can be easily measured from finger using a single PPG sensor. Therefore, PPG signal has been widely used for BP estimation due to its simplicity and easy-to-use as well as cheapness. Besides, PPG's derivatives contains much information helpful to BP estimation, and this technique has been widely used in related studies

(Yang et al. 2020a; Baek et al. 2019; Slapničar et al. 2019; Cheng et al. 2021; Harfiya et al. 2021; Wang and Zhang 2017; Liu et al. 2021; Attarpour et al. 2019; El-Hajj and Kyriacou 2021b; Lin et al. 2021a; Shimazaki et al. 2018; Baek et al. 2020; Xing et al. 2019; Atomi et al. 2017). Specifically, the first-order differential PPG signal, also known as velocity plethysmography (VPG), contains slope information related to BP. The second-order differential PPG signal, also known as accelerated plethysmograph (APG), contains dominating information about the dichroic notch and diastolic point (Cheng et al. 2021).

There are two other variants of PPG technique in addition to single site PPG. The first is *multi-wave PPG signals* (MWPPG) which is acquired based on different colors of light. For example, Baek et al. (2020) measured single multi-wave PPG signals using a smartphone for BP prediction, and finds that the best performance is achieved when a green PPG signal is used in conjunction with an instantaneous frequency signal. Liu et al. (2020a) proposed a PCA-based MWPPG algorithm for BP prediction using only a single sensing node, where MWPPG decodes the compounded multi-wave PPG signals into arterial pulse and capillary pulse, and the phase lag between them is used further to compute PTT. The second is *multi-channel PPG signals* (MCPPG) which is acquired at different sites (Attarpour et al. 2019; Fong et al. 2019; Lazazzera et al. 2019). For example, Fong et al. (2019) proposed a SVR-based ensemble method for BP prediction using MCPPG signals collected from multiple arterial segments of an individual's left arm, where each SVR in the ensemble is trained on a comprehensive feature set that is derived from a distinct PPG signal.

ECG signal-based ECG signal-based BP estimation has attracted some attention recently since ECG signals are easy to collect using wearable devices. Based on literature search, we find that current studies related are all based on a single lead ECG signal (Fan et al. 2019, 2021; Landry et al. 2019; Wu et al. 2016; Miao et al. 2020; Haddad et al. 2021; Mousavi et al. 2020; Simjanoska et al. 2020; Wu et al. 2016). The fusion of multi-lead ECG signals for BP estimation may be a potential research direction.

Piezoelectric signal-based Piezoelectric (PZT) sensor can sense pressure changes and convert them into electrical signal. Therefore, it is suitable for arterial distension sensing during the cardiac cycle (Samartkit et al. 2022; Yi et al. 2022b). PZT sensor is usually attached on the subject's wrist through a wrist strap and can work without an external power source, making it superior to other high-power sensors while providing safety insurance (Park et al. 2017; Samartkit et al. 2022). Wang and Lin (2020) proposed to use PZT signal for beat-by-beat BP monitoring, where the pressure change is converted from the voltage change by the pressure sensitivity of the sensor. However, the relation between PZT signal and BP waveform remains unclear. Recently, Yi et al. (2022a) elucidated the first derivative correlation between PZT signal and BP waveform for the first time based on theoretical analysis and experimental simulation, which lays foundations for BP monitoring using a single PZT signal.

OMW-based Oscillometric waveform (OMW) signal represents a class of oscillography. In traditional oscillography methods, SBP and DBP that usually obtained by mapping the position at predetermined ratio of the envelope of the maximum amplitude of the OMW signal to the deflation curve are very sensitive to the ratio. Therefore, several researchers treat OMW as regular signal source for BP estimation under regression scenario where ML and DL methods are adopted to solve the nonlinear relationship between time domain features extracted from OMW signal and the reference BP (Lee et al. 2018; Lee and Lee 2020; Lee and Chang 2016, 2017a, b, 2019; Forouzanfar et al. 2011; Lee et al. 2019a, 2020). Besides, Argha et al. (2019); Argha and Celler (2019) firstly model OMW-based BP prediction as a sequence-to-sequence classification question. Specifically, each OMW beat is labeled with one of three classes, namely pre-systolic (PS), between systolic and diastolic

(BSD) and after diastolic (AD). Then, SBP and DBP points are determined based on the beat at which the model output sequence switches from PS to BSD and from BSD to AD, respectively.

AW-based Auscultatory waveform (AW) represents a class of auscultation methods. Traditional auscultation methods require the participation of experts. Recently, some researchers are committed to using advanced machine learning technology to realize automatic BP estimation based on AWs (Celler et al. 2019a; Argha et al. 2020; Pan et al. 2019). For example, Celler et al. (2019a) proposed a Gaussian mixture Models and hidden Markov model (GMM-HMM) method to automatically discover and learn the latent structure in the AW sequence, and then SBP and DBP points are determined as the cuff pressures at which the AW sequence changes its structure. Similar to previous studies (Argha et al. 2019; Argha and Celler 2019), Argha et al. (2020) models AW-based BP estimation as a sequence-to-sequence classification question.

Multi-sensor signal-based Multi-sensor signal fusion (Khaleghi et al. 2013) technologies have been widely used in BP prediction area. Specifically, for traditional explicit-feature-extraction-based methods, the features extracted from multiple synchronized signals are concatenated together to form the final feature vector, which is usually followed by feature selection/reduction methods to reduce dimension before it is used for training. While, for DL methods with raw signal as input, multiple signals are combined in some form (refer Sect. 5.2.2) and then are fed into neural network for model training.

Since the amount of information about BP collected by a single sensor is often limited, and the signals collected by different sensors (usually worn it on different sites of the body, with different working principle) have certain complementarity, modeling all kinds of influence factors about BP through the fusion of multi-sensor signals is promising. A large number of studies have confirmed that the BP prediction accuracy of the method based on multi-sensor signal fusion is better than that of any method based on a single sensor signal (Lee et al. 2021; Paviglianiti et al. 2020a; Esmaelpoor et al. 2021b; Thambiraj et al. 2020; Huang et al. 2022; Baek et al. 2019). ECG signal contains important information about BP, which can effectively improve the prediction accuracy as a supplement to PPG signal (Esmaelpoor et al. 2021a). It is observed that PPG and ECG signals are the most popular combination used for BP prediction. Due to its simple configuration and portability, PZT may be a promising alternative to ECG for calculating pulse transit time together with PPG in conventional PTT methods (Samartkit et al. 2022). Besides, other signals such as PCG, BCG and ICG are used in conjunction with PPG and/or ECG signals for better BP prediction (Samartkit et al. 2022). In short, BP prediction based on multi-sensor signal fusion has potential applications prospects in wearable devices. However, this inevitably increases the difficulty of configuration and proposed new challenges for storage and response time when deployed on hardware.

3.4.2 Health Behavior data based

One of the advantages of health behavior data-based BP monitoring is that the convenience to explore the primary factors affecting individual's BP, which provides an opportunity for health behavior recommendations. Chiang and Dey (2018) investigated the relationship between several health behavior such as sleep and exercise and daily BP based on the proposed random forest with feature selection (RFFS). Experimental results show that the healthy behavior recommendation function of RFFS can in turn be used to regulate

Table 13 Categorization of different physiological signal source for BP prediction

Input signal		Studies
PCG		Peng et al. (2015)
PZT		Yi et al. (2022a), Park et al. (2017), Wang and Lin (2020)
OMW		Lee and Chang (2016, 2017a), Lee et al. (2018, 2019a, 2020), Lee and Lee (2020), Lee and Chang (2019), Argha and Celler (2019), Forouzanfar et al. (2011)
AWs		Celler et al. (2019a), Argha et al. (2020), Pan et al. (2019)
SCG		Das et al. (2020)
PPG	Single PPG	Rong and Li (2021a), El Attaoui et al. (2020), Esmaelpoor et al. (2020), Wang et al. (2017),Ibtehaz and Rahman (2020), Monte-Moreno (2011), Riaz et al. (2019), Chen et al. (2022),Schlesinger et al. (2020, 2020), Yan et al. (2019), Lin et al. (2021a), Kurylyak et al. (2013),Xing and Sun (2016), Liu et al. (2017), Bose and Kandaswamy (2018); Bose and Kandaswamy (2017), Chowdhury et al. (2020),Schrumpf et al. (2020, 2021, 2021), Gao et al. (2016) , Dey et al. (2018), Lin et al. (2021b),Mousavi et al. (2019b), El-Hajj and Kyriacou (2021b, 2021), Hsu et al. (2020), Kido et al. (2022),Khalid et al. (2020), Leitner et al. (2021), Hasanzadeh et al. (2019), Li et al. (2021),Panwar et al. (2020), Dagamseh et al. (2021), Sasso et al. (2020)
	MCPPG	Attarpour et al. (2019), Ibrahim and Jafari (2019), Fong et al. (2019), Fujita et al. (2019),Brophy et al. (2021), Cheng et al. (2021), Sadrawi et al. (2020), Lazazzera et al. (2019),Athaya and Choi (2021), Aguirre et al. (2021), Harfiya et al. (2021), Slapničar et al. (2019),Dörr et al. (2021), Li and He (2021), Attarpour et al. (2019), Maher et al. (2021)
	MWPPG	Baek et al. (2020), Liu et al. (2020a)
ECG	SIECG	Fan et al. (2019, 2021), Landry et al. (2019), Wu et al. (2016), Simjanoska et al. (2020),Haddad et al. (2021), Mousavi et al. (2020), Wu et al. (2016), Miao et al. (2020)
PPG+	ECG	Kachuee et al. (2016), Sharifi et al. (2019), Senturk et al. (2020), Esmaelpoor et al. (2021b),Liu et al. (2021), Su et al. (2018), Song et al. (2021, 2019), Qiu et al. (2021), Hill et al. (2021),Paviglianiti et al. (2020a), Wang et al. (2021), Thambiraj et al. (2020), Yang et al. (2021),Huang et al. (2022), Zhang et al. (2017), Chuang et al. (2021), Yamanaka et al. (2021),Yang et al. (2020a), Miao et al. (2017), He et al. (2016a), Liu et al. (2020b), Yin et al. (2021),Li et al. (2020a) Baek et al. (2019), Mahmud et al. (2022), Zhang et al. (2021a), Ji et al. (2022),Jeong and Lim (2021), Esmaelpoor et al. (2021a), Franco et al. (2019), Baker et al. (2021),Lo et al. (2017), Tanveer and Hasan (2019)
AWs +	OMW	Celler et al. (2019b)
ECG+	OMW	Ahmad et al. (2012)
ECG+	PPWs	Miao et al. (2019)
PPG+	PCG	Marzorati et al. (2020)
PPG+	BCG	Yousefian et al. (2020)
PPG+	PZT	Samartkit et al. (2022)
PPG+	BCG+ECG	Zhang et al. (2021c), Eom et al. (2020), Lee et al. (2021)

Table 13 (continued)

Input signal		Studies
PPGs+	ECG+SCG	Ganti et al. (2020).
PPG+	ECG+ICG	Sola et al. (2013)
PPG+	ECG+PCG	Esmaili et al. (2017)

MCPPG multi channel PPG, *MWPPG* multi wavelength PPG, *SlECG* single lead ECG, *PPW* pulse pressure wave, *OMW* oscillometric waveform, *AWs* auscultatory waveforms, *SCG* seismocardiogram, *PCG* phonocardiogram, *BCG* ballistocardiogram, *ICG* impedance-cardiogram, PZT piezoelectric

individual's blood pressure. Chiang et al. (2021) further proposed a method called RFSV for BP prediction and personalized recommendation. Specifically, the best ARIMA model's parameter is used to extend the original feature set, and then a general random forest (RF) model is trained using the whole feature set. Next, RF with shapley value (RFSV) technique is utilized to select important features to train model for final prediction and recommendation.

3.4.3 Trajectory data based

Trajectory data based BP monitoring is a potential way for unobtrusive BP monitoring, since no wearable device is needed, and trajectory data can be easily acquired from mobile devices. Currently, there are few relevant studies. Xiang et al. (2021) firstly proposed a framework for BP prediction using individual's daily trajectory data in conjunction with demographical characteristics. As Fig. 20 illustrates, through grid-based clustering algorithm, the trajectory data is firstly converted to region-of-interest (ROI) label sequences, which is followed by Bayesian topic model comprising the LDA for acquiring the probability of daily routine pattern distribution. Finally, the acquired daily routine topic distribution, historical BP and demographical features are used as input to train a LSTM model for prediction.

3.4.4 Facial video based

Facial video based methods are a class of contactless, cuffless methods that only use facial video for BP estimation, which overcome the drawback that PPG-based methods are sensitive to contact pressure (Chandrasekhar et al. 2020) and have a very wide range of potential application values.

For video-based methods, an important technique is how to obtain the PPG component (aka iPPG signal) from video. Based on Lambert-Beer law and light scattering theory, a general method in firstly determine ROI from video, and then PPG component is extracted from ROI of the video based on spatial averaging operation (Tasli et al. 2014; Sugita et al. 2015; Jeong and Finkelstein 2016; Secerbegovic et al. 2016; Fan et al. 2018; Takahashi et al. 2020; Djeldjli et al. 2021; Zhou et al. 2019; Luo et al. 2019).

The development of video-based BP measurement has experienced a process similar to that of physiological signal-based BP measurement. Specifically, facial video-based BP prediction methods can be generally grouped into three categories, namely (i) mathematical/optical methods, (ii) video-based ML methods, (iii) video-based DL methods, etc.

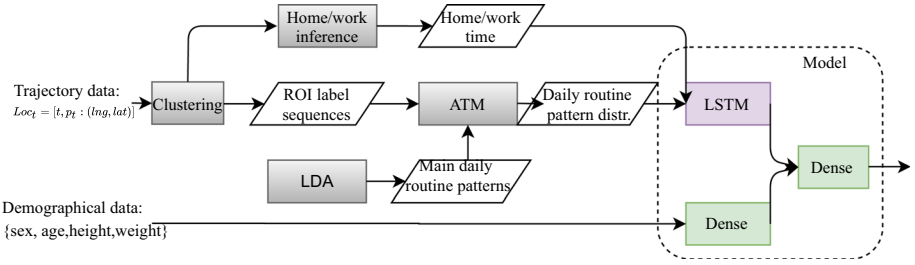

Fig. 20 Model framework proposed in study (Xiang et al. 2021)

Mathematical/optical methods includes image-based PTT (iPTT) methods, etc. Image-based PTT (iPTT) methods (Sugita et al. 2015; Jeong and Finkelstein 2016; Secerbegovic et al. 2016; Fan et al. 2018) is similar to traditional PTT methods. In iPTT methods, the pulse transient time or instantaneous phase difference is firstly calculated between two or multiple pulse waves obtained from different parts (such as face and palm) of an individual's body captured by a video camera. However, subjects are often required to maintain a fixed posture, which is unrealistic in real-life scenarios. Hence, Takahashi et al. (2020) examined the feasibility of acquire PTT based on a single part of the body, and the strong correlation between PTT acquired from the forehead and the chin of a face and BP is confirmed. In addition, Zhou et al. (2019) proposed a multiple channel averaging-based methods for BP estimation using RGB camera. Specifically, the facial video is firstly converted from RGB space to YCrCb space for face detection. Spacial averaging is then performed for each channel of the ROI image, and the obtained average is used as input to the JADE algorithm for blind source separation. Based on radial resonance theory, the peaks and valleys of the resulting time series signal are related to pressure waves.

For video-based ML methods (Djeldjli et al. 2021; Luo et al. 2019; Rong and Li 2021b; Gonzalez Viejo et al. 2018), after PPG component is extracted from video, relevant features are extracted from the PPG component, which is further used for training BP prediction model. For example, Djeldjli et al. (2021) proposed a single channel averaging-based methods to acquire iPPG signal. Specifically, after ROI is detected, the green channel of the ROI image is selected and averaged to acquire iPPG signal, which is further used for the extraction of BP-related features. Rong and Li (2021b) acquired three iPPG signals from the ROI of the channels of videos, and a total of 26 features are extracted from these signals, Then, four algorithms including LR, SVR, RF and MLP are employed to train BP prediction model based on these features, respectively. In the well-known transdermal optical imaging (TOI) method (Luo et al. 2019), subtle changes in facial image is captured to to detect blood pulsation in cardiovascular system, which is further used for BP estimation. Specifically, 17 ROIs of facial image with robust hemoglobin fluctuations are selected. Then averaging operation in performed in each ROI, resulting in 17 hemoglobin signals. Next, features are extracted from these signals to train a MLP model for BP prediction.

For video-based DL methods, although we have not found any ready-made literature, please note that video-based technologies for other related tasks such as heart rate monitoring have achieved some progress (Yu et al. 2019). In addition, Yu et al. (2019) firstly use a deep spatio-temporal network for reconstructing precise remote PPG (rPPG) signal from facial videos. We believe that video-based DL methods will also usher a new period of vigorous development as in physiological signal-based methods in the near future.

4 Datasets and evaluation

In this section, we plan to summarize the widely used public datasets and processing tools, followed by an overview and commentary on the widely used evaluation metrics, evaluation procedure, and splitting strategies. Finally, we made a critical analysis of the reported results.

4.1 Dataset and processing toolbox

Since the publication of the first Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) dataset in 1996 by Moody and Mark (1996), there have been several freely

accessible databases supporting BP estimation. In this subsection, we summarized several representative datasets in this field and the relevant supporting tools for data processing.

Table 14 summarized several popular datasets used in related studies. It can be seen that several datasets such as MIMIC xxx and PPGBP are constantly expanding, producing sever different versions. The specific version of the dataset used should be declared when using a dataset. Especially note that datasets such as MIMIC xxx, etc., are heterogeneous dataset collected from ICU. The collected signals are accompanied with various noise interference and measurement error, etc., it can not be directly used for experiment. Therefore, Kachuee et al. (2015) published a pre-processed version of the MIMIC II dataset named UCI-BP, this version of the dataset has been widely used in this field due to its ease-of-use. However, note that the sampling rate of ECG signal is reduced from 500 to 125 Hz to keep synchronization with PPG signal, which may lead to time jitter of about 8 ms in the extraction of PTT-related features (in the worst case). This small problem has been neglected in most of the studies, and Sharifi et al. (2019) proposed to tackle this problem by averaging PTT based on cubic splines.

Besides, several researchers use virtual database for experiments. For example, Huttunen et al. (2019) used a database of virtual subjects generated based on 1D haemodynamic model where model parameters are varied to reflect variations between different subjects. Magbool et al. (2021) used two publicly available, virtual pre-validated databases of simulated pulse waves for experiments.

Table 15 presents several widely used toolboxes for processing and analyzing physiological signals.

4.2 Evaluation metrics

The metrics used in classification scenario include accuracy, precision, recall, and F1 Score, etc. The metrics used in other learning scenarios include mean absolute error (MAE), mean absolute percentage error (MAPE), mean square error (MSE), mean error (ME), standard error (STD), and R^2 (Monte-Moreno 2011), etc. Herein, we focus mainly on the latter, and a summary of these metrics is in Table 16. Especially, MAE is related to the BHS standard (O'Brien et al. 1993), ME and STD are related to the AAMI standard. Therefore, these three metrics are widely used for performance comparison in related studies. However, as Fig. 21 illustrates, *a single metric MAE is insufficient to objectively evaluate the performance of an algorithm on a dataset with skewed distribution*. This is inspired by the fact that the generalization ability of a model may be very poor even if the *Accuracy* value is high in the class-imbalance scenario. That is why the existence of metrics such as F1 Score, etc. Naturally, *is there such a metric like F1 Score in regression scenario?*

Based on the above considerations, we introduced a new evaluation metric named bin-balanced MAE (b^2MAE), and propose the MAE of prediction for each bin corresponding to different BP levels should also be reported, in addition to the global MAE. The definition of b^2MAE is as follows.

$$b^2\text{MAE} = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \frac{1}{|\{y_j \in bin_i\}|} \sum_{y_j \in bin_i} |y_j - \hat{y}_j|, \tag{15}$$

Table 14 Several popular datasets for blood pressure prediction[a]

Dataset	Individual state	Collection location	Description	Waveform			Sampling frequency	Publication source	References
				PPG	ECG	ABP			
MIMIC I Goldberger et al. (2000), Moody and Mark (1996)	ICU patients	Boston's Beth Israel Hospital	Contains over 90 subjects with totally 121 records	✓	✓	✓	PPG, 125 ECG, 500 ABP, 125	Computers in Cardiology, 1996	Wang et al. (2018b), Fati et al. (2021), Tanveer and Hasan (2019), Paviglianiti et al. (2020b), Zhang et al. (2021a), Jeong and Lim (2021), Ji et al. (2022), Chuang et al. (2021), Kurylyak et al. (2013)
MIMIC II (v2.0 ~ v2.6) Saeed et al. (2011)	ICU patients	Boston's Beth Israel Hospital	The latest version (v2.6) contains 32536 subjects with totally 40426 admissions	✓	✓	✓	PPG, 125 ECG, 500 ABP, 125	Crit. Care Med.,2011	Maher et al. (2021), Qiu et al. (2021), Xing and Sun (2016), Esmaelpoor et al. (2020), Schlesinger et al. (2020), El Hajj and Kyriacou (2020a), Schlesinger et al. (2020), Khalid et al. (2020), Liu et al. (2017), Hassani and Foruzan (2019), Senturk et al. (2020), Landry et al. (2019), Fan et al. (2021), Esmaelpoor et al. (2021a), El-Hajj and Kyriacou (2021a), Tazarv and Levorato (2021), He et al. (2016a)

Table 14 (continued)

Dataset	Individual state	Collection location	Description	Waveform			Sampling frequency	Publication source	References
				PPG	ECG	ABP			
MIMIC III (v1.0∼ v1.4) Johnson et al. (2016)[b]	ICU patients	Boston's Beth Israel Hospital	The latest version (v1.4) contains approximately sixty thousand admissions of patients	✓	✓	✓	PPG, 125 ECG, 500 ABP, 125	Sci. Data, 2016	Miao et al. (2020), Baker et al. (2021), Schrumpf et al. (2021b), Yang et al. (2020a), Leitner et al. (2021), Fan et al. (2019), Chen et al. (2019), Slapničar et al. (2019), Aguirre et al. (2021), Lin et al. (2021b), Hill et al. (2021)
MIMIC IV (v0.4 ∼ v1.0) Johnson et al. (2021)	ICU patients	Boston's Beth Israel Hospital	An update to MIMIC-III, which incorporates contemporary data	✓	✓	✓	PPG, 125 ECG, 500 ABP, 125	PhysioNet	–
QUVD Liu et al. (2012)	Anesthesia state	Royal Adelaide Hospital	32 cases ranging in duration from 13 min to 5 h	✓	✓	×	100	Tech. Commun., 2012	Huang et al. (2022), Zhang and Wang (2017), Khalid et al. (2020, 2018), Brophy et al. (2021), Tazarv and Levorato (2021)
PPGBP (v1 ∼ v3) Liang et al. (2018)	Outpatients with hypertension and diabetes	Guilin People's Hospital	219 subject records, each lasting 6.3 seconds	✓	×	×	1000	Sci. Data, 2018	Chowdhury et al. (2020), Tjahjadi et al. (2020), El Attaoui et al. (2020), Maqsood et al. (2022)

Table 14 (continued)

Dataset	Individual state	Collection location	Description	Waveform			Sampling frequency	Publication source	References
				PPG	ECG	ABP			
UCI-BP Kachuee et al. (2015)	ICU patients	Boston's Beth Israel Hospital	12000 subject records	✓	✓	✓	125	ISCAS, 2015	Kachuee et al. (2016), Sharifi et al. (2019), Baek et al. (2019), Schlesinger et al. (2020), Thambiraj et al. (2020), Kachuee et al. (2015), Mousavi et al. (2019b), Rong and Li (2021a), El-Hajj and Kyriacou (2021b), Huang et al. (2022), Wang et al. (2020), Kachuee et al. (2015), Hasanzadeh et al. (2019), Panwar et al. (2020), Thambiraj et al. (2019), Li et al. (2020a), Schrumpf et al. (2021b), Ibtehaz and Rahman (2020), Brophy et al. (2021), Cheng et al. (2021), Li and He (2021), Harfiya et al. (2021), Sharifi et al. (2019), Hsu et al. (2020), Bose and Kandaswamy (2018), Bose and Kandaswamy (2017), Franco et al. (2019), Li et al. (2021), Lo et al. (2017)
VitalDB Lee and Jung (2018)	Surgical patients	Seoul National University Hospital	Time-synchronized biosignals of 6388 patients	✓	✓	✓	100	Sci. Data, 2018	Lee et al. (2019b), Yang et al. (2020a), Zhang et al. (2021a)
Esmaili et al. (2017)	Healthy individuals	Sharif University of Technology	32 subjects	✓	✓	×	1000	IEEE TIM, 2017	Esmaili et al. (2017)

[a] *MIMIC* Medical Information Mart for Intensive Care, *QUVD* the University of Queensland Vital signals Dataset, *ICU* intensive care units, *PPG* plethysmograph, *ECG* electrocardiogram, *ABP* arterial blood pressure, *SpO2* blood oxygen saturation, *PCG* phonocardiogram, *FSR* force-sensing resistor

[b] MIMIC III is an update to MIMIC II v3.2

Table 15 Toolboxes for processing and analyzing physiological signals

Toolbox	Description
Neurokit2 (Makowski et al. (2021)	A user-friendly Python package for simulating and processing various neurophysiological signals such as PPG, ECG, RSP, EDA, and EMG, etc.
Scipy Virtanen et al. (2020)	A well-known Python package including modules for statistical, optimization, integration, Fourier transform, etc. Its sub-module Scipy. signal implements various filters and peak detection, interpolation and transformation algorithms for signal
WFDB Vijayarangan et al. (2020)	A well-known, source software available for all popular platforms, which supports signal processing, automatic analysis, visualization, annotation and interactive analysis of waveform
wfdb-python Vilalta and Drissi (2002)	A Python package for reading, writing and processing WFDB signals (e.g ECG) and annotations. This package provides a python interface to access MIMIC series databases
Peak detection Virtanen et al. (2020)	A Python package including peak detection algorithm
Vayu Mahajan (2021)	A open-source, cross-platform sensor data analysis toolbox, including data con- version, interpolation, aggregation and prediction

Table 16 Summary of metrics used in the evaluation of BP estimation

Metric	Calculation formula	Description
MAE	$\frac{1}{N}\sum_{i=1}^{N}\lvert y_i-\hat{y}_i\rvert$	Calculate the MAE of prediction over all samples
MAPE	$\frac{1}{N}\sum_{i=1}^{N}\frac{\lvert y_i-\hat{y}_i\rvert}{y_i}$	Calculate the MAPE of prediction over all samples
MSE	$\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$	Calculate the MSE of prediction overall sample, sensitive to abnormal predictions
ME	$\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)$	Calculate the ME of prediction over all samples
STD	$\sqrt{\frac{1}{N}(y_i-\hat{y}_i)^2}$	Calculate the STD of prediction over all samples, the square of metric MSE
R^2	$1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2}$	Measure the fitting effect of the model, the closer its value is to 1, the better

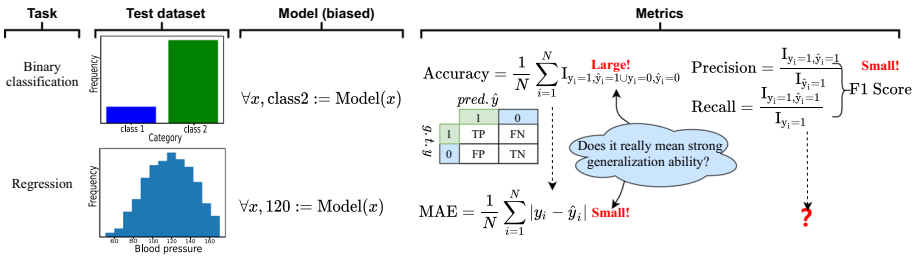

Fig. 21 In class imbalance scenario, a single metric-Accuracy can not objectively evaluate the performance of classifier. Similarly, a single metric-MAE is insufficient to evaluate the performance of a model in imbalanced regression scenario

where N_{bin} refers the number of bins with equal length s included in the total BP range $[bp, \overline{bp}]$, i.e. $N_{bin} = \lceil (\overline{bp} - bp)/s \rceil$, parameter s controls the granularity. The i-th bin $\overline{bin_i} = (bp + (i-1)s, bp + i \cdot \overline{s}]$.

It is intuitive that the 'Mean' operation is performed within each bin to offset the imbalanced distribution. Similarly, bin-balanced version of other conventional metrics can also be easily derived.

4.3 Evaluation procedure and splitting strategy

The selection of evaluation procedure is dependent on the learning scenario. Specifically, for *online/incremental learning scenario*, the usually used evaluation procedure is prequential evaluation (Gama et al. 2009), i.e each sample in the data stream is firstly served as test sample for prediction, and then is used as training sample to update prediction model. For *sequence prediction scenario*, the usually used evaluation procedure is sequential test, i.e the former part of the sequence data is used to train model and the latter part of which is used for test. For *other learning scenarios*, the usually used evaluation procedures include: (1) cross evaluation, i.e the whole dataset is split into multiple equal portions (e.g. 10 portions, which is called 10-fold cross-validation). In each iteration, a portion of data is used to test the model trained on the remaining portions of data, and the loop does not terminate until each portion of data has been used as test set; (2) random splitting, i.e the whole dataset is randomly split into training, validation, and test set according to a certain ratio, the experiment is repeated by changing random seed to obtain multiple different divisions. However, *the splitting operation in the above two evaluation procedures itself is a critical factor that affect the final division, which is usually ignored in most of the literature*.

Specifically, as Fig. 22a and b illustrates, the splitting can be performed at the record level or sample level. For splitting strategy (a), all samples of each record appears only in training, validation or test set. Whereas, this strategy may leads to large differences of BP distribution among training, validation and test set, especially when the total number of records is small. Therefore, careful check is required to ensure the BP distribution among the three sets is consistent (Schrumpf et al. 2021a, b). However, there are only a few studies where the consistency check is performed and disclosed (Schrumpf et al. 2021a, b; Song et al. 2019; Yang et al. 2020a; Atomi et al. 2017; Bose and Kandaswamy 2018). For example, Bose et al. confirmed the consistency between training and test sets in terms of several characteristics such as SBP, DBP, Body mass index (BMI), etc. Besides, we noticed that in several studies (Qin et al. 2021), all BP records are firstly divided into several disjoint subsets based on BP category, each subset is then split into training, validation and test records, the final splitted dataset is acquired by merging the records with same kind from different subsets.

For splitting strategy (b), the final aggregated samples are randomly divided to form training, validation and test sets. In other words, samples of a record may appears simultaneously in training, validation and test sets. However, physiological signal of an individual is highly regular and will not change significantly in a short time, especially, signals are usually collected from individuals in an enclosed environment (we means participants' range of activity, status and posture when measuring data), leading to insufficient BP variations. Therefore, this splitting strategy is at the risk of data leakage, although there is no intersection among training, validation and test sets(Eom et al. 2020; Schrumpf et al. 2021a, b; Hasanzadeh et al. 2019).

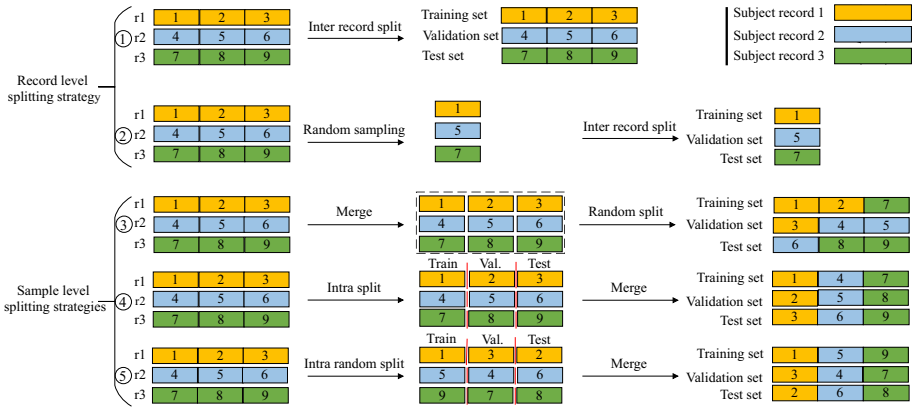

Fig. 22 Several splitting strategies used for evaluation. Splitting strategy—(**a**) ensures all samples of a record appear only in training, validation or test set, refer rows ①∼② for details. Nevertheless, for splitting strategy—(**b**), samples of a record may distributed among training, validation and test set, refer rows ③∼⑤ for details. **c** denotes a special experimental protocal that sample level splitting strategy is used for experiment, and experiments are performed individual-by-individual, refer ⑥∼⑦ for details

In addition, there are other experimental procedures such as Leave-one-subject/record-out (LOSO) (Yu et al. 2019), and individual test (Yan et al. 2019; Liu et al. 2020b; Lin et al. 2021b), etc. LOSO can be regarded as a special case of cross-validation with record level splitting strategy, in which only one record data is used as the test set in each iteration. Since BP changes vary greatly among different individuals (Zhang et al. 2021c), the test results on different individuals often vary widely. In individual test (as illustrated in Fig. 22c), the experiment is performed individual-by-individual. Specifically, for each individual, part of the individual's data is used to train a personalized prediction model, and the remaining data is used to further test the model. Since the limited individual BP dynamics and the small amount of data, the trained model is subject-specific and its generalization ability is very limited, although experimental results usually seems well.

4.3.1 Quantitative comparison

To quantitatively analyze the effect of different splitting strategies-① &③ &④ &⑤ (refer Fig. 22. Note that splitting strategies-② &⑥ &⑦ were excluded considering the too small number of samples included or unbearable computational cost) on experimental results, we utilize the classical ResNet (He et al. 2016b) and MIMIC III database for our experiments. Concretely, ResNet was modified for BP prediction (2D convolution was replaced by 1D convolution, and the last classification layer was replaced by regression layer consisting of two neurons). For dataset, we used the version published by Schrumpf et al. (2021a) and randomly select 750 records (i.e v_1 version) for experiments. The statistics of several datasets finally used are summarized in Table 17 and a graphical illustration of the BP distribution and individual BP dynamics is attached in Appendix 2. It can be seen that both SBP and DBP cover an extensively large range in terms of overall population BP and individual BP dynamics.

Table 18 presents the global numerical results. Moreover, considering the severely skewed BP distribution both in training set and validation set, the corresponding test

Table 17 Statistics with respect to total BP as well as individual BP dynamics, of the final used datasets derived from MIMIC III

Dataset	Type	#Data	Total population			Individual dynamics		
			Range	Mean	Std	Range	Mean	Std
v0	SBP	375 records, totally 7.5e5 samples	50–199	122.23	24.62	58–145	104.70	15.27
	DBP		40–119	61.75	12.60	32–78	58.01	11.65
v1, (default)	SBP	750 records, totally 1.5e6 samples	50–200	122.59	24.22	58–145	104.80	15.22
	DBP		40–119	61.68	12.41	25–78	58.22	11.45
v2	SBP	1500 records, totally 3e6 samples	44–200	122.44	24.33	58–147	105.27	14.74
	DBP		40–119	62.03	12.65	25–78	58.92	11.49
v3	SBP	2250 records, totally 4.5e6 samples	40–200	122.40	24.23	58–147	105.62	14.51
	DBP		40–119	61.96	12.52	25–79	59.02	11.50
v4	SBP	3000 records, totally 6e6 samples	40–200	122.26	24.33	58–153	105.67	14.85
	DBP		40–119	61.83	12.48	25–79	58.91	11.70

v0⊂v1⊂v2⊂v3⊂v4

performance of different models on each bin of test range is visualized in Appendix 3 (Figs. 28, 29). We can find several interesting observations as follows,

(1) The test result of ResNet model based on sample level splitting strategies-③ &④ &⑤ is significantly better than that based on record level splitting strategy-①. However, note that this does not necessarily ensures the former with stronger generalization ability, since data of an individual appears simultaneously in training set and validation set;
(2) No matter which splitting strategy is adopted, the trained model prefer to make predictions towards central BP region. In other words, skewed data sets lead to biased model. As illustrated in Figs. 28 and 29, the corresponding MAE of the model increase gradually as the test bin is farther away from the central region of the possible BP range, which is inversely proportional to the BP distribution of training set. This circumstance is even more serious when the record level splitting strategy is used. One may question whether this phenomenon is caused by too few training samples away from the central area. Our experimental results on more large datasets such as v_4 (4x larger that v_1) indicate that this phenomenon is still occurred.
(3) No matter which splitting strategy is adopted, although the BP distribution among training, validation and test set is confirmed to be consistent, the performance of the model on validation set and test set is lower than that on training set by a large margin, which is even more obvious when record level splitting strategy is used. We confirm that no over-fitting phenomenon is occurred during training, and believe that this is mainly caused by individual differences, which challenges the training of general models with robustness, and strong generalization ability.

In addition, Friedman test (Demšar 2006) is employed to judge whether the performance of ResNet models based on different splitting strategies is comparable. Specifically, the mean rank of each method over five experiments is calculated and the resulting p-value is 3.57e−3 (0.05) for both SBP and DBP prediction. Therefore, the null hypothesis is rejected at $\alpha = 0.05$, i.e splitting strategy has a significant effect on the trained model's performance for both SBP and DBP prediction. Furtherly, post-hoc Nimenyi test (Nemenyi

Table 18 Comparison of the performance of ResNet models based on different splitting strategies for BP prediction on MIMIC III dataset consisting of totally 15e4 samples from 750 records. Splitting ratio is set to 6:2:2. The presented results are the average of the five experimental results under the random splitting (refer Subsection 4.3) experimental protocol

Splitting strategy	–	SBP						DBP					
		MAE	MAPE	MSE	ME	STD	R^2	MAE	MAPE	MSE	ME	STD	R^2
①	Training set	2.787	0.023	13.623	− 0.346	3.687	0.976	1.999	0.033	6.859	− 0.132	2.614	0.956
	Validation set	12.103	0.102	316.704	− 0.276	17.791	0.462	6.683	0.110	95.266	− 0.019	9.759	3.748
	Test set	12.288	0.103	330.257	− 0.280	18.170	0.456	6.571	0.107	91.823	0.070	9.579	0.396
③	Training set	2.891	0.024	14.485	0.055	3.804	0.975	2.041	0.034	7.070	0.074	2.657	0.954
	Validation set	8.196	0.069	159.348	0.178	12.623	0.728	4.713	0.076	50.598	0.227	7.113	0.671
	Test set	8.188	0.069	158.869	0.179	12.604	0.729	4.712	0.076	50.647	0.22	7.117	0.671
④	Training set	2.923	0.024	14.810	− 0.204	3.848	0.975	2.112	0.035	7.551	− 0.034	2.747	0.951
	Validation set	8.393	7.046	164.481	− 0.093	12.825	0.721	4.855	0.079	52.760	0.163	7.263	0.658
	Test set	8.403	7.052	164.402	− 0.092	12.822	0.719	4.851	0.079	52.481	0.134	7.244	0.658
⑤	Training set	2.948	0.025	14.923	− 0.288	3.862	0.975	2.086	0.035	7.360	− 0.101	2.712	0.952
	Validation set	8.406	0.071	164.740	− 0.171	12.835	0.720	4.838	0.079	52.549	0.075	7.249	0.659
	Test set	8.404	0.071	164.473	− 0.159	12.825	0.720	4.830	0.078	52.292	0.078	7.231	0.660

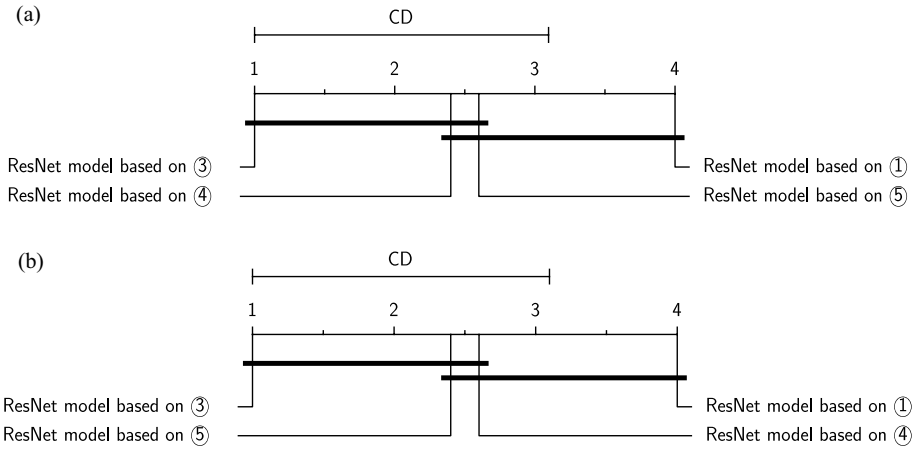

Fig. 23 Visualization of post-hoc Nimenyi test (Nemenyi 1963) (in terms of MAE) of ResNet models based on different splitting strategies, for **a** SBP prediction, **b** DBP prediction

1963) is applied for pairwise performance comparison. Specifically, a critical difference at 95% confidence interval is 2.0976, and the results are visualized in Fig. 23. It can be seen that ResNet model based on record level splitting strategy-① achieves the worst ranking among models based on other splitting strategies, and ResNet model based on sample level splitting strategy-③ achieves the best ranking among models based on other splitting strategies. The p-value for ResNet models based on splitting strategies-① &③ is 1.363e−3 for both SBP and DBP prediction. Therefore, the null hypothesis is rejected at $\alpha = 0.05$, and we conclude splitting strategy is statistically significantly factor relating to model's test result. Moreover, we find that only models based on splitting strategies-③ &④ are statistically comparable when the number of experiments exceeds twelve since the rank of models in each experiment is relatively stable.

4.4 Analysis of reported results

Instead of simply comparing the results as in other studies, in Table 22, we conducted a comprehensive comparison of the latest/representative literature on BP prediction based on twelve metrics. We can find several basic facts as follows,

(1) The prediction results reported in different articles vary greatly, even for those that using the same data source and similar methods. This naturally brings people some confusion when comparing the results of related articles: First, *what makes the results reported in different articles so different*? Second, *for those papers reporting good results, where does the performance improvement come from? data cleaning, feature engineering, algorithm improvement or hyper-parameter optimization*?
(2) The performance of models trained on different datasets using the same method varies greatly.
(3) The prediction accuracy reported in some studies is obviously unrealistic (note that we are not questioning the authenticity of the experiment, but the procedural irrationality).

(4) The accuracy reported in those studies based on ICU patients' data is generally lower than that based on data collected from healthy or ordinary individuals.
(5) Those studies with fewer records/subjects included for experiment are more likely to produce better results. Since the experimental settings vary from studies to studies, a more rigorous experiment (refer Appendix 5) by us more fully confirmed this point.
(6) The inclusion of demographic characteristics for building model can improve prediction accuracy.
(7) Calibration/fine-tuning techniques can help to significantly improve the prediction accuracy of the model.
(8) The reported accuracy in those studies based on individual test or experiments with sample level splitting strategy is usually significantly higher than those based on experiments with record/subject level splitting strategy.

We would like to state a basic view-*a single good result does not necessarily ensure the model with strong generalization ability*. For example, due to the limited and stable variation of individual BP, the personalized model usually performs very well under individual test protocal. However, the performance of the model will be significantly degraded when it is not calibrated for a long time or the physiological activity of the individual changes, let alone on other individuals that have never been seen during training. That is why intra-individual BP variation is so important for the evaluation of individualized estimation methods with individual-by-individual calibration (Liu et al. 2020a). In fact, Mukkamala et al. (2021) has disclosed that the conclusions of an increasing number of publications are potentially misleading. Herein, we summarized possible reasons for unfairness in comparisons and some of the possible reasons for unreliable results, mainly from the perspective of an ML researchers.

Possible reasons for unfairness in comparison. Unlike other ML/DL application areas such as computer vision where there are plenty of baseline methods and out-of-the-box public datasets. In the BP estimation community, due to the lack of baseline methods and ready-to-use datasets (note that although several public datasets have been summarized in Sect. 4.1, there are still many processing procedures to be done before it can be used in experiments, such as data cleaning, signal denoising, segmentation, data splitting, etc.), the direct comparison method (i.e comparing a system) is widely used for comparison. However, this method of comparison is unfair due to the following reasons,

(1) Even if the same data source is used, the final processed data sets used in the experiment may vary greatly, since the sample duration, the signal preprocessing method and the data cleaning procedures vary from studies to studies (Schrumpf et al. 2021a, b; Slapničar et al. 2019; Tazarv and Levorato 2021; Mousavi et al. 2019b).
(2) The splitting ratio used for the generation of training, validation, and test sets varies from studies to studies.
(3) The experiment procedure as well as splitting strategy used vary from studies to studies, which is one of the main reasons for significant differences between the results reported in different studies.

Possible reasons for the unreliability of some reported results. We summarized several factors accounting for the unreliability of some results from an ML researcher's perspective, as follows,

(1) A large proportion of papers are based on private data that is non-accessible. In addition, for those literatures based on public data, the final processed dataset is rarely made public and usually can not be exactly regenerated due to missing or inadequate description of the details such as parameter configuration.
(2) The data set used in some studies contains too few records with narrow BP range and insufficient/sparse BP variations, which means the conclusions itself has great limitations. In fact, He et al. (2016a) has experimentally confirmed that the performance of predictor degrades significantly when it is applied to individuals with dramatically fluctuating BP values. More seriously, the BP range of test set is narrower than that of training set in some studies, which is problematic.
(3) Some articles did not even disclose the details (the number of records/samples, statistics on the range of BP and its distribution, in the final processed dataset) of the data used.
(4) The data leakage issue both in the process of normalization and feature selection/reduction, test set should not be included in the above two processes.
(5) There are few studies where the consistency check of the BP distribution between training set and test set (the characteristic of I.I.D) is performed and reported, which is an critically important procedure for the objective assessment of ML models.
(6) The widely used splitting strategy at the final sample level is at the risk of data leakage, resulting in some unrealistic results, which further leads to overestimation of the performance of model, although there is no sample overlap between training, validation and test set.
(7) There are some studies where a single metric is used to evaluate the performance of the model.
(8) There are a few literatures where the *normalize target* operation was used during training, and the reported evaluation results were computed based on the normalized predictions.
(9) Due to the severely skewed BP distribution in the data set, metric values such as MAE that evaluated on the whole data set are not sufficient to explain the performance of the model.
(10) The currently used evaluation standards for traditional BP measurement methods seems no longer necessarily suitable to ML/DL based methods (Mukkamala et al. 2021).

It is observed that the factors leading to unreliable results involve almost all links of the blood pressure prediction pipeline, many of which are more or less related to the data itself. It is worth mentioning that, in addition to contributing to the unreliability of the results, some of the above-mentioned factors also bring about a reproducibility crisis to a certain extent. Besides, in this area, there are few authors publish code along with their papers (Slapničar et al. 2019). Here, we list all the open-source implementations we can find for BP prediction in Table 20 (Appendix 4).

5 Some critical issues and techniques

In this section, we will discuss and summary several key issues as well as newly-emerging techniques in the BP prediction community in the form of special topics.

5.1 Critical issues

On the basis of a large number of literature analysis, we condensed six issues to be discussed, which are *imbalanced phenomenon*, *interpretability issue*, *sample duration*, *individual difference*, *large difference between SBP and DBP prediction accuracy*, and *handcrafted features versus machine-learned features*.

5.1.1 Imbalanced phenomenon

In a large number of related studies (Kachuee et al. 2016; Radha et al. 2019; Schrumpf et al. 2021a; Jeong et al. 2019; Dagamseh et al. 2021; Esmaelpoor et al. 2020; Schrumpf et al. 2020; Schlesinger et al. 2020; Sagirova et al. 2021), when analyzing the distribution of MAE for different SBP and DBP values, there is a similar expression such as "for the bins far from the central area within the total BP range, the larger the corresponding MAE". Actually, considering the scatter plot, as shown in Fig. 24, the slope of the fitted line from the (ground-truth, prediction) pairs is always smaller that 1 (ideally, the slope should be equal to 1). In other words, the BP of samples in the area with relatively low BP is overestimated, and the BP of samples in the area with relatively high BP is underestimated.

Herein, we formally declare that this is the effect of imbalance phenomenon. From a statistical point of view, the usually used MSE loss in regression modeling is equivalent to the negative log likelihood loss of a noisy prediction distribution-$p(y|x;\theta)$, of which the mean is the model's prediction (Bishop and Nasrabadi 2006). We further assume the label-conditional distribution-$p(x|y)$ is the same in both training and test set, then by Bayes's rule, it is easily derived,

$$\frac{p_{\text{train}}(y|x)}{p_{\text{test}}(y|x)} \propto \frac{p_{\text{train}}(y)}{p_{\text{test}}(y)}, \tag{16}$$

Equation (16) indicates that the ratio between $p_{\text{train}}(y|x)$ and $p_{\text{test}}(y|x)$ is proportional to $p_{\text{train}}(y)$, which is lower when a BP value rarely appears in the training set. Summarily, the BP dataset with skew/imbalanced distribution leads the trained model to make prediction biased towards central BP region.

Generally speaking, imbalanced regression is a novel and challenging topic in the whole machine learning community(Krawczyk 2016). Currently, the imbalance issue in BP prediction has hardly been paid attention to in related studies. So far, we have found several tricks appeared in a few studies (Radha et al. 2019; Tjahjadi et al. 2020; Wang et al. 2022)

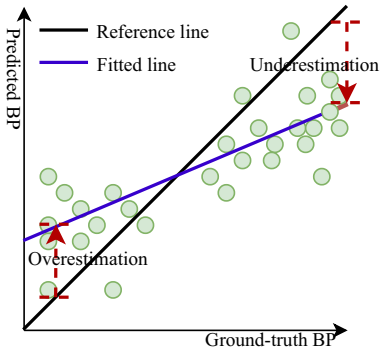

Fig. 24 A demo interpreting a general case of scatter plot of blood pressure prediction results reported in related studies

which actually serve to mitigate the imbalance issue, although this is not explicitly stated in the text. Radha et al. (2019) believe that there will be a problem when the commonly used mean square loss that favours minimizing samples with large errors is used for the BP prediction task in which the genuine BP presents a normal distribution. Therefore, a strategy that amplifying the loss of samples by the absolute difference of the corresponding BP from the mean BP is proposed, which is similar to cost sensitive learning in classification scenarios. Tjahjadi et al. (2020) used up-sampling technique to overcome the imbalance issue between different BP groups. Wang et al. (2022) proposed a modified loss function based on quantile, in which the loss of the samples whose genuine BP is lower that quartile 1 of the overall BP in the population or higher than quartile 3 of the overall ground-truth BP in the population is magnified.

5.1.2 Interpretability issue

The lack of interpretability and poor robustness are two important, common issues of artificial intelligence technologies when applying it into specific application fields. Specifically, predictive models/systems should be explainable to understand how they work and the predictions should be realistic and consistent with basic principles, which is crucial for adopters to be confident when using them to aid decision making (Sethi et al. 2020; Moss et al. 2022). This issue has been mentioned in Sect. 3.2.3. Generally, the higher the model complexity, the lower the interpretability. However, the complexity of healthcare decisions often requires the use of complex models. Therefore, for a specific task, there is a trade-off between model performance and model interpretability (Moss et al. 2022; Sethi et al. 2020). Here, we plan to conduct a more in-depth analysis in conjunction with relevant latest research.

Currently, there are little related work focuses on the interpretability of AI models for BP prediction. For studies applying traditional feature-based ML methods, in addition to using interpretable ML algorithms such as decision tree and gradient-boosting tree (Zhang et al. 2019a) to ensure interpretability, the focus is on eliciting interpretable features such as PIR (Ding et al. 2017; Ding and Zhang 2015), Womersley number (Thambiraj et al. 2020, 2019) and contact pressure (Chandrasekhar et al. 2020), which can be used to both clinically justify an non-invasive BP estimation and inspire new research on physiological correlates of BP. Since current AI is intrinsically data-driven (Zhang et al. 2020a), for those studies applying DL models, how to combine it with related prior knowledge/correlations (such as those discussed in Sect. 2) to infer more biologically and physically realistic models for robust estimation is a challenging while meaning topic. For example, Kissas et al. (2020) proposed a physics-informed neural network for BP estimation where data-driven DL models was seamlessly synthesized with one-dimensional blood flow model derived from first physical principles for the first time, and the final model can return physically consistent predictions. Besides, there are many post hoc methods (Moss et al. 2022) to achieve model interpretability, such as Shapley values (SHAP), Locally interpretable model-agnotic explanations (LIME), etc. These methods are model-agnostic, and therefore can be applied to various models.

5.1.3 Sample duration

For signal data, to acquire a dataset available for training, raw signal has to be divided into disjoint signal segments, each of which corresponds to a sample. There are two

types of segmentation methods: *i*) segment based on fixed *time duration*; *ii*) non-uniform segmentation.

The commonly used method is segment based on fixed time duration. Time duration is usually set to 5 s(Slapničar et al. 2019), 7 s (El-Hajj and Kyriacou 2021a), 10 s (Fan et al. 2021), 15 s (Mousavi et al. 2019b), 30 s (Schlesinger et al. 2020), 60 s (Monte-Moreno 2011) in related studies. Time duration, as a hyperparameter, affects experimental results, and its effect has been investigated in a few studies (Schrumpf et al. 2021a, b; Simjanoska et al. 2020). Schrumpf et al. (2021a, b) experimentally finds that different time duration (1, 2, 5, 7, 9, 11, 13, 15, 17, 20 s) resulted in an almost equal prediction error, and the maximum time duration of 20 s is recommended for experiment. Simjanoska et al. (2020) performed experiments using ECG signals with different time duration (10, 20, 30 s), filtered with different cut-off frequencies (0.05~0.5 Hz), and confirmed that for the *final sample level* splitting strategy, time duration of 30 s filtered with cut-off frequency of 0.35 Hz leads to the best result. While, for the *record level* splitting strategy, time interval of 10 s filtered with cut-off frequency of 0.30 Hz produces the best result. Sasso et al. (2020) performed experiments using different sample duration (15, 30, 45 s) both in the stress test data and 24-h data, and finds that sample duration of 30 s leads to relatively better results in average.

However, this segment approach will cause an interruption of PPG cycles at the beginning and the end of each segment. Moreover, the resulting varying number of cycles in each segment will induces bias in the extracted features towards subjects with higher heart rate (Tanveer and Hasan 2019).

Non-uniform segmentation is proposed to overcome the above mentioned issues by dividing according to fixed number of cycles. In this context, the so-called beat-by-beat BP prediction (Miao et al. 2020; Esmaelpoor et al. 2020; Xing and Sun 2016; Bose and Kandaswamy 2018; Singla et al. 2020a) can be viewed as a special case of non-uniform segmentation where each beat/cycle of signal segment corresponding to a sample. However, please note that since deep learning model with raw signal as input accepts only the input of fixed length, the non-uniform segments have to be resampled (e.g. zero-padding or signal interpolation (Yang et al. 2020a; Li et al. 2021; Dey et al. 2018), etc.) to make its length the same.

For example, Tanveer and Hasan (2019) proposed a non-uniform waveform segmentation method where PPG segment with a length of three consecutive systolic peaks as well as ECG segment with a length of three R-peaks are extracted from raw signals, then the normalized two segments are resampled to fixed length and then are concatenated to form the final waveform-based feature vector. Schrumpf et al. (2021a, b) divided both PPG and ABP signals into segments containing distinct fixed number of cycles, and then these segments were resampled to have equal length. Experimental results indicate that the prediction errors based on non-uniform segmentation are lower compared to that based on fixed time duration.

5.1.4 Individual difference

The existence of individual differences is a significant characteristic of physiological signal data, which increased the difficulty of learning tasks. Taking BP prediction as an example, the relationship between input signal and BP may vary from individual to individual, since each person had unique and subtly different cardiovascular dynamics (Zhang et al. 2021a). A more extreme example is that the same PPG cycle shapes do not always guarantee the

same BP values (Slapničar et al. 2018). There are a total of three strategies to overcome individual difference in this area.

The first is based on *divide and conquer strategy*. Khalid et al. (2018) experimentally confirmed the significant prediction error differences among normotensive, hypertensive and hypotensive groups, and suggests that the future BP prediction model should be more specific for different BP categories. Generally, the total population is divided into several disjoint groups based on one or more characteristics, such as age and BP category, etc., and then each group of data is used to train a distinctive prediction model. In the test phase, an additionally trained classifier is firstly used to predict the belonging group, and then the corresponding model is called for prediction. Actually, individual test (Miao et al. 2019) can be seen as a special case where each individual is treated as a distinct group. There are several studies following this schema (Yamanaka et al. 2021; Maher et al. 2021; Khalid et al. 2020; Chen et al. 2022). For example, Dey et al. (2018) proposed an ensemble of six Lasso regression models for BP prediciton where each model is trained on a distinct group of data determined based on the value of age, gender, and BMI.

Instead of training multiple models, Simjanoska et al. (2018) directly use the predicted group of the classifier to extends the feature set which is further used for training prediction model. Further, Simjanoska et al. (2020) proposed an ensemble of the three multi-target regression models trained on each BP group for BP prediction where the ensemble weights are determined by the output probabilities of the additional classifier.

The second is the inclusion of demographical features (aka personal information), since demographical features such as age and BMI are critical influencing factors related to individual BP state (Yang et al. 2020b). This technique has been widely used in related studies to improve BP prediction accuracy. Concretely, for traditional feature-based methods (Song et al. 2021; Monte-Moreno 2011; Attarpour et al. 2019; Yin et al. 2021; Atomi et al. 2017; Shimazaki et al. 2018; Dey et al. 2018; Datta et al. 2016; Yamanaka et al. 2021; Zhang et al. 2019a; Liu et al. 2021; Simjanoska et al. 2018; Chowdhury et al. 2020), demographical features are directly used to extends the feature set. For deep learning methods with raw signal as input (Liu et al. 2018; Xiang et al. 2021; Koshimizu et al. 2020; Yang et al. 2021; Lee et al. 2021), demographical features are usually embedded to the last layers of the neural network model via a fully-connected layer.

The third is the utilization of domain adversarial training technology, which is usually in conjunction with with deep learning. Specifically, in addition to the learning task, an additional classifier is introduced to enforce the model to learn cross-individual features by adversarial training. In this scenario, the optimization target can formulated as,

$$\max_{\theta_d} \min_{\theta_g, \theta_f} L_r(f(g(x;\theta_g), \theta_f), y) - \lambda L_d(d(g(x;\theta_g), \theta_d), l_x), \tag{17}$$

where L_r, L_d denote the main task loss and classifier loss. θ_f, θ_g and θ_d denote the parameters of feature learner, task network, classifier module, respectively. Experimental results indicate that domain adversarial training technique can boost model training, enables the predictive model with better generalization ability to other individuals, allowing less target domain samples for training accurate personalized model (Zhang et al. 2020b; Qin et al. 2021).

In addition, individual-by-individual BP centralization (aka zero meanization) technique is used in several studies (Yang et al. 2020a; Miao et al. 2020; Haddad et al. 2021). In other words, model is trained to prediction BP variation instead of genuine BP. Specifically, BP data of each individual in the training set is subtracted by its mean value during training,

and in the test phase, for each individual, the final predicted BP is the sum of the model's prediction and its mean BP, which can be seen as one-time calibration.

5.1.5 Large difference between SBP and DBP prediction accuracy

Based on extensive literature analysis, we find an interesting phenomenon-*the prediction accuracy of SBP reported is significantly lower than that of DBP* (Jeong et al. 2019; Miao et al. 2020; Das et al. 2020; Haddad et al. 2021; Esmaelpoor et al. 2021b; Mousavi et al. 2019b; Attarpour et al. 2019; Thambiraj et al. 2020; Rong and Li 2021a; Wang et al. 2021; Qiu et al. 2021; El-Hajj and Kyriacou 2021b; Yin et al. 2021; Huang et al. 2022; Lin et al. 2021a; Xing and Sun 2016; Bose and Kandaswamy 2018; Bose and Kandaswamy 2017; Zhang et al. 2019b; Baker et al. 2021; Fong et al. 2019; Esmaelpoor et al. 2020; Wang et al. 2020; Zhang and Wang 2017; Wang and Zhang 2017; Singla et al. 2019; Baek et al. 2020; El Hajj and Kyriacou 2020a; Chiang and Dey 2018; Datta et al. 2016; Kachuee et al. 2015; Dastjerdi et al. 2017; Liu et al. 2018; Schlesinger et al. 2020; Schrumpf et al. 2021a, b; Yamanaka et al. 2021; Liu et al. 2020b; Baek et al. 2019; Zhang et al. 2019a; Chiang and Dey 2019; Yousefian et al. 2020; Liu et al. 2021; Li and Laleg-Kirati 2021; Zhang et al. 2021c; Chen et al. 2021; Miao et al. 2019; Yang et al. 2020a; Liu et al. 2020a; Leitner et al. 2021; Chiang et al. 2021; Hasanzadeh et al. 2019; Ibrahim and Jafari 2019; Kachuee et al. 2016; Fan et al. 2019; Esmaili et al. 2017; Slapničar et al. 2018; Simjanoska et al. 2020; Wang et al. 2018b; Zhang et al. 2020b; Yang et al. 2021; Radha et al. 2019; Thambiraj et al. 2019; Simjanoska et al. 2018; Chen et al. 2019; Slapničar et al. 2019; Eom et al. 2020; Chowdhury et al. 2020; Li et al. 2020a; Lee et al. 2021; Fati et al. 2021; Li et al. 2021). This phenomenon is consistent with physiological explanations. Concretely, the relation of DBP and its variability with aortic stiffness are generally weaker than those of SBP and its variability, and the popular datasets in this area (refer Sect. 4.1) are mainly collected from ICU patients, surgery patients and outpatients suffering from a variety of cardiovascular diseases. However, there are a few literatures (Simjanoska et al. 2018; Khan Mamun and Alouani 2022) where the reported SBP prediction accuracy is unexpectedly higher than that of DBP. We argue this abnormal phenomenon is originates from the special statistical characteristics (we refer to the BP range and BP distribution) of the collected dataset.

Musini and Wright (2009) confirmed that the coefficient of variation of SBP was significantly greater than the coefficient of variation of DBP. In fact, based on the statistical description about the data used in related studies and in our study (refer Table 17 and Fig. 27), the range and standard deviation of SBP in the population are usually significantly larger than those of DBP in the population, which undoubtedly increases the difficulty of SBP prediction. This is no problem in traditional STL scenario. As mentioned in Sect. 3.3.2, however, this is an important issue to consider when designing MTL model for BP prediction.

5.1.6 Hand-crafted features versus machine-learned features

It is generally assumed that hand-crafted features are limited since it can not adequately express the information in the input signals related to BP variations. On the other hand, DL enables automatic feature learning from raw signals, making it more and more popular in cuffless BP estimation community. However, it is unclear whether there is a difference

between these two types of features and which type is better. Currently, few articles pay attention to this problem.

Mahmud et al. (2022) proposed a novel MLP model for BP prediction, in which a pre-trained U-Net model is used as feature learner. Experiments show that a U-Net model trained by mapping PPG to ABP instead of PPG signal can help the MLP model achieve the best performance. Its high predictive accuracy is impressive. Shimazaki et al. (2018) found in the experiment that the combination of hand-crafted features and learned features based on auto-encoder enables the predictive model to obtain better performance than the model based on any single type of features. Esmaelpoor et al. (2021a) compares the effects of physiological parameter and learned feature based on CNN network in BP prediction, and they fond that the learned features are superior over physiological parameters, and the combination of these two types of features does not improve prediction performance. What's embarrassing is that the conclusions in the two articles are exactly the opposite. It should be noted that the conclusion in study (Esmaelpoor et al. 2021a) is limited since the extracted features are relatively few, and the learned features are based solely on CNN network which is trained by predicting BP. We recommend other types of network can be tried and the network used for outputting learned features can be trained with diverse purposes (such as reconstructing input signal, predicting BP, etc.)

5.2 Techniques

The techniques to be discussed include *data augmentation*, and *different signal combination schemes*.

5.2.1 Data augmentation

Data augmentation is an important technology to solve the problem of insufficient data in training complex models, especially deep neural networks. We noticed that, in computer vision area, there are many popular data augmentation techniques (Shorten and Khoshgoftaar 2019; Hussain et al. 2017) are available for image data. However, it is unclear what data augmentation techniques are applicable to signal data, and there is no relevant systematic review, which is what we are going to discuss here. We summarized three types of data augmentation techniques for signal data used in BP prediction research.

The first is cropping-based data augmentation technique. Esmaelpoor et al. (2020) used cropping technique for enhancing the training set. Specifically, for each PPG segment, additional ten sub-segments are cropped for better describing the time-domain possible relationships.

The second is filter-based technique. Huang et al. (2022) proposed a novel multi-filter to multi-channel (MFMC) technique to generate multi-channel signal to adapt the input format of the MLP-mixer architecture. The so-called MFMC technique is that the multi-channel signals are derived by applying multiple distinct filters and filtering parameters to the original signal. Simjanoska et al. (2020) tried to generate multiple datasets/configurations for final ensemble learning by setting different cut-off frequency and sample duration. Specifically, by setting sample duration of 10, 20, 30 s and cut-off frequencies starting from 0.05 Hz up to 0.50 Hz by step of 0.05 Hz, a total of 3×10 datasets are generated.

The third is parametric Bootstrap method (Lee and Chang 2016, 2017a; Lee et al. 2018, 2019a, 2020). In parametric Bootstrap method, mean and standard deviation are firstly estimated using limited training set based on Normal distribution assumption (note that this

process is performed feature by feature), then the bootstrap samples based on the estimated distribution were calculated using the Monte Carlo method. Furtherly, Song et al. (2021) proposed a similar parametric Bootstrap method based on multivariate Gaussian distribution (MGD) where the relationship between the features is incorporated in a multi-dimensional feature vector.

5.2.2 Different signal combination schemes

In BP estimation area, techniques such as fusion of multiple different signals (e.g PPG and ECG signals, etc. (Kachuee et al. 2016; Miao et al. 2017; Baek et al. 2019; Miao et al. 2019; Yang et al. 2020a; Song et al. 2019; Baker et al. 2021)), multi-channel homogeneous signal (e.g multi-channel PPG signals, etc. Attarpour et al. 2019; Fong et al. 2019; Lazazzera et al. 2019), multi-wave signals (Baek et al. 2020; Liu et al. 2020a), different modalities of homogeneous signal (generated based on multi-order difference and time domain/time-frequency domain transformation Baek et al. 2019; Slapničar et al. 2019; El-Hajj and Kyriacou 2021b; Rong and Li 2021a; Harfiya et al. 2021) are usually used to improve BP estimation accuracy. Naturally, a practical issue in *how to effectively combine these input signals for deep learning methods with raw signal as input*?

Currently, there are three signal combination schemes. The first is to directly concatenate different signal segments in temporal direction (Shimazaki et al. 2018; Tanveer and Hasan 2019). For example, Tanveer and Hasan (2019) directly concatenate PPG and ECG segments in temporal direction, which is fed into an ANN-LSTM network for training. What we want to emphasize is that this combination scheme is limited to specific network, which may be problematic when it is used in convolutional network, since different signals have different varying patterns and temporal dynamics.

The second is to concatenate different signal segments in channel direction. This is reasonable since the feature map of each channel in a tensor represents distinct patterns extracted from the input. The network, correspondingly, is built based on 1D convolution, which is widely used in related studies (Baek et al. 2019; Cheng et al. 2021), since different signals are synchronous signal.

The third is to consider different signals through multi-branch structure (Slapničar et al. 2019; Rong and Li 2021a; Baek et al. 2019, 2020). Intuitively, different signals are processed using independent network module separately, and then the outputs of which are further fused based on FFNN module.

In addition, Qiu et al. (2021) proposed a 2D-convolution-based network for BP prediction where two PPG and ECG segments are stacked and treated as a picture for processing. However, although the two signals are intrinsically related, signals collected from multi-modal sensors have modality-specific characteristics. Shared weights for whole input signals may lead to interference between features, which originates from capturing modality-specific features of different signals. Ha and Choi (2016) proposed to extracts modality-specific features and common features for human activity recognition task based on partial-weight sharing and full weight sharing strategy, which may inspire us to devise more effective network models with multi-modality signals as input, for BP prediction.

6 Some related machine learning topics

In this section, we introduce several advanced ML technologies for potential applications in BP estimation.

6.1 Auto ML/AI

Although there has made much progress in the application of ML and DL for BP prediction, humans are heavily involved in almost all aspects (such as feature engineering, the selection of feature selection method, model selection, etc.) of constructing ML prediction models. In fact, Chowdhury et al. (2020) has investigated the best combination of feature selection methods and training algorithms manually in order to improve BP prediction accuracy, but it is very cumbersome.

AutoML (Waring et al. 2020; He et al. 2021) is a new-emerging technology for automatically building a specialized system without human assistance. Although AutoML has made significant progress, it is rarely mentioned in the field of healthcare monitoring such as BP estimation (Waring et al. 2020). Fati et al. (2021) firstly proposed a BP prediction model based on the tree-based pipeline optimization tool (TPOT), which automatically selects the best combination of training algorithm and feature selection method from the library for SBP and DBP estimation, separately.

In addition to TPOT, there are also other AutoML pipelines, such as H2O (LeDell and Poirier 2020), Auto-sklearn (Feurer et al. 2019) and FLAML [304]. Especially, for DL methods, there is neural architecture search (NAS) (Elsken et al. 2019) which automatically finds the best neural network architecture configuration for any given dataset and learning task.

6.2 Transfer learning

Transfer learning (Pan and Yang 2009) is novel framework aiming at improve the learning of target domain using related source domain. Taking BP prediction into account, a single individual contains relatively little data, which is insufficient (consider data amount and BP variations) to train a robust model with strong generalization ability. Therefore, *is it possible to consider leveraging other individual's data in some way to facilitate training*? On the other hand, there are plenty of freely released pretrained deep learning models especially in CV areas, *is it possible to leverage these models for BP estimation*? Following the above two questions, related works were categorized into two folds, as Fig. 25 illustrates.

Leitner et al. (2021) formally proposed a transfer learning framework for BP prediction based on convolutional-recurrent neural network (CRNN) for the first time. Ablation studies indicate that the best performance is acquired when finetune the specific layers (the last Conv. and FC layers) of CRNN during knowledge transfer. Final experiments demonstrated that the performance of the finetuned model is superior over the general model (i.e no calibration) and the model trained from scratch (i.e using only limited target data). Nevertheless, we would like to emphasize that transfer learning is not a novel topic in BP estimation area, although it is not explicitly mentioned in the relevant literature. Actually, there have been a large count of studies where both record/subject level splitting strategy and calibration technique have been used for experiment (Leitner et al. 2021; Kachuee et al. 2016; Slapničar et al. 2019; Schrumpf et al. 2021a, b; Song et al. 2019; Qin et al. 2021; Zhang et al. 2020b;

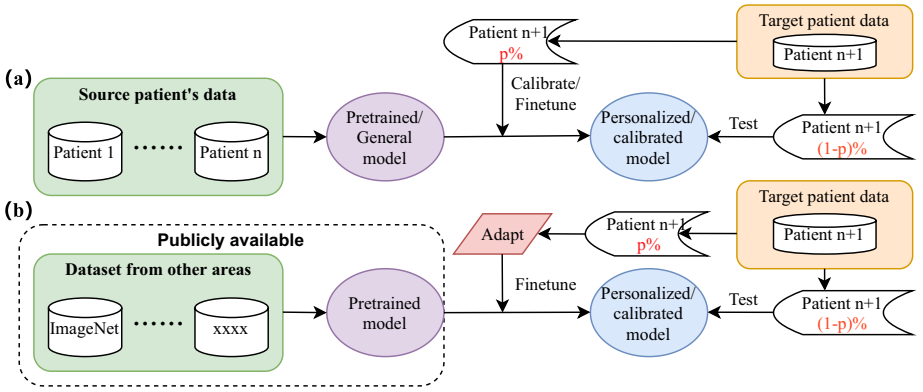

Fig. 25 Two representative transfer learning scenarios for blood pressure prediction. **a** other patients data is served as source domain for knowledge transfer; **b** pretrained models from other areas such as computer vision are directly used for knowledge transfer

Bose and Kandaswamy 2018; Kachuee et al. 2015). Since the existence of individual difference, the transferred model usually performs poor on the test individual, and calibration procedure is employed to calibrate the model to adapt to the test individual using limited data from test individual. From this point of view, *subject/record level experiment + calibration* $\subseteq$ *transfer learning*. Besides, there are a few authors first train a pretrained model using PPG-BP dataset, then a personalized model is fine-tuned from the pretrained model using rPPG-BP dataset (Schrumpf et al. 2020, 2021a, b). In the above studies, all of the source domain individuals are used to train a general model for further knowledge transfer. However, since different individuals have different BP levels and BP dynamics, more intelligent selection of source individuals suitable for target individual for knowledge transfer may help to improve prediction accuracy (Zhang et al. 2020b; Leitner et al. 2021).

When employing pretrained CV models for knowledge transfer, the input signal has to be processed to adapt to the input format of the model, and the model needs to be modified to perform BP estimation. Sasso et al. (2020) proposed a method by fine-tuning the pretrained ResNet-18 model on ImageNet using HYPE dataset. Specifically, the time-domain PPG segments are firstly converted to image representations (spectograms and scalograms), which are then fed into ResNet-18 for model fine-tuning. Wang et al. (2020) proposed a method by fine-tuning the pretrained Inception V3 model on ImageNet using MIMIC II database. Specifically, the last FC layer with softmax activation is replaced with FC layer contains two neurons and with linear activation. Next, time-domain PPG segment is firstly converted to image based on the visibility graph technique and then the self-replicated images are fed into Inception V3, and only the FC layer is updated for BP prediction during fine-tuning.

6.3 Meta learning

Meta learning (Vilalta and Drissi 2002; Vanschoren 2018) is a novel learning framework that learn meta-knowledge from a variety of tasks, such that it can generalize well on new task when using only limited data from the new task. Taking BP estimation into account, with similar motivation to transfer learning, meta learning attempts to learn how to learn from other individual's data and then quickly generalize to test individual by fine-tuning using

limited test data. Intuitively, the experimental protocol is under *subject/record level experiment + calibration*. Currently, there are few studies in this area. Cheng et al. (2021) proposed a convolutional network with U-Net architecture for ABP waveform reconstruction where the well-known model-agnostic meta-learning (MAML) algorithm is utilized for model training. Specifically, each record is treated as a learning task, the model is firstly initialized using the pre-training set. Next, a learning task is randomly selected in each iteration, support set and query set are acquired from the corresponding record, support set is used to update task parameters and generate the task model. Query set is then used to evaluate the personalized model and update the global model. Finally, for any test individual, the global model is fine-tuned using limited data from the individual, and the remaining data is used for test.

6.4 Federated learning

Federated learning (Li et al. 2020b) is new-emerging technology that tackles data sharing and privacy issues by training a global model over remote devices, such as mobile phones, while keeping data localized. Considering the importance of user privacy protection and data security, this technology has great development potential in healthcare area (Hakak et al. 2020). For example, Brophy et al. (2021) proposed a CycleGAN-based model for generating ABP waveform from PPG sigal under the federated learning framework. Specifically, to simulate the decentralized environment, the whole dataset is split into multiple disjoint parts, each of which represents a terminal. A localized model is trained on each terminal, and then these models are send to a global model for aggregation. Next, the aggregated model can be used to perform downstream task or used to update each localized model.

6.5 On-device machine learning

On-device ML (Dhar et al. 2021) is a technology that running (including model training, model inference) machine learning on edge devices, which propose new challenges to the requirements for model size and time delay, due to the limited resources such as memory and computing power. Currently, a variety of smartphone-based health applications are emerging, due to the widespread popularity of smartphones with high-resolution cameras and built-in sensors such as accelerometers, orientation sensors. We noticed that there are several smartphone-based BP monitoring literatures (Matsumura et al. 2018; Chandrasekaran et al. 2012; Dey et al. 2018; Visvanathan et al. 2014; Luo et al. 2019; Sagirova et al. 2021) where the models used are either explicit analytical models based on hemodynamics, or traditional ML models such as LR, SVM, MLP, etc. Although DL for BP prediction has been extensively studied at the academic level, as far as we know, there is no application that has been successfully deployed on devices such as smartphones and has been certified by relevant institutions. We think this is a potential and meaningful research direction.

7 Discussions and conclusions

In this section, we discuss the challenges in BP prediction, and the question of *what a good BP estimator should look like?*, which is followed by a general proposal towards the objective evaluation of model's performance. Finally, we end this survey with conclusions and several potential research directions.

7.1 Challenges

Non-invasive BP prediction is a meaningful while challenging issue in healthcare monitoring area. Specifically, the challenges include the follow three aspects,

(1) Complexity of problem/data itself. Firstly, as mentioned in Sect. 3.2.1, the signal data collected by the sensor is seriously disturbed by all kinds of noise, which seriously affects the quality of the data. Secondly, during data collection, the health status of the participants may be different (such as diabetes, obesity and other cardiovascular diseases), and the measurement status may be various (rest or exercise, standard or sit, alcohol/drug intake, mood, etc.), all these factors more or less affect the collected data and indirectly affect the prediction accuracy of the model, which undoubtedly increases the difficulty of the problem. Since these factors have different characteristics from the measured data (Li et al. 2017), it is worthwhile to further investigate how to effectively take these factors into account when training the model and how to quantitatively assess the impact of these factors on the model. Thirdly, the existence of individual differences is a significant characteristic of physiological signal dataset, which means the underlying relationship between input and BP may vary from individual to individual. Even, there may be situations where two individuals have similar input signals but the measured BP differs significantly (Slapničar et al. 2018). Resultly, there may be some other types of 'mismatch' between training set and test set in addition to the overall BP range and BP distribution.
(2) The severely skewed/imbalanced distribution of BP in dataset (we call *target imbalance*, which is similar to class imbalance in classification scenario, but is more challenging). As noted in Sect. 5.1.1, imbalanced regression is a challenging and neglected problem in the ML community and almost all related application fields including BP prediction, although this phenomenon has been done and mentioned several times in related studies. Moreover, target imbalance is usually intertwined with other imbalance factors such as imbalance between the number of samples among different individuals (we call *record/subject imbalance*) (Khalid et al. 2018; Schrumpf et al. 2021a), etc. For record/subject imbalance, a common practice is to control the number of samples from different individuals to a specified number (Schrumpf et al. 2021a, b).
(3) How to objectively evaluate the performance of BP prediction algorithms is an opening and challenging questions in this area. As mentioned in Sects. 4.3 and 4.2, we have disclosed some critical factors responsible for the objective evaluation of BP prediction model from multi-aspects such as data itself, evaluation strategy, evaluation metrics, etc.

7.2 A BP estimator with good generalization ability

The current progress of BP prediction in summarized in Table 22. The experimental results reported in different papers vary greatly, even for the same data source. We further disclosed several factors that lead to this phenomenon and some unreasonable practices across through the experiments in Sect. 4.3.1. It seems that a single result does not necessarily reflect the generalization ability of the model. Herein, we discuss the question of *what a BP estimator with good generalization ability should look like?*

(1) *Good performance over different BP intervals* The results (e.g MAE, etc.) reported in almost all related studies are average results across the total BP range bounded by the BP of test set. However, as detailed in Sect. 4.2 , this is insufficient because the poor performance of the biased model on the regions far from the central BP range is covered up by the severely skewed distribution of BP data set. However, overestimating the BP of individuals with hypotension, especially underestimating the BP of individuals with hypertension, will seriously mislead doctors' decision-making and may cause irreparable losses. Therefore, a good BP estimator should perform well in different BP intervals within the possible BP range.
(2) *Good performance over different individuals* It seems that significant results have been achieved in BP prediction, especially those studies based on sample level splitting strategies or individual test schemes. However, please note that the goal is to train a general BP estimator by using the data of limited individuals to make it generalize well on "unseen" individuals. A good BP estimator should perform well on different individuals, especially those individuals never appeared during training and validation processes.
(3) *Good performance over different databases*. A good BP estimator should perform well on different BP data sets. However, currently, there are only a few studies (Miao et al. 2020; Xing and Sun 2016; Yang et al. 2020a; Huang et al. 2022) where external validation is performed.

7.3 A general proposal-towards objective evaluation of model's performance

Mukkamala et al. (2021) argue that the increasing number of papers on BP prediction that pass traditional evaluation criteria are methodologically inadequate and misleading, and further revealed the capabilities and limitations of these methods based on several solid experiments. It seems that passing conventional evaluation standards (such as BHS O'Brien et al. 1993, AAMI Zhang et al. 2020b, etc.) and analysis tools (such as Bland-Altman plot, etc.) may not necessarily guarantee good performance.

Corresponding to Sects. 4.4 and 7.2, we *give an overall proposal* by examining the entire pipeline shown in Fig. 6 in order to objectively evaluate the performance of predictive models.

(1) Both the training set and the test set should contain enough BP variations as well as diversity in terms of age, etc., which is necessary to train a general model and for objective evaluation. Commonly used means of changing BP include exercise (such as rope skipping, running, etc.), cold stimuli and brain activity (e.g mental arithmetic), etc (Lin et al. 2020; Esmaili et al. 2017; Miao et al. 2017; Block et al. 2020; Ding et al. 2017; Ibrahim and Jafari 2019; Ganti et al. 2021). Moreover, necessary check must be made to ensure that the BP distribution between the training set and the test set is consistent, which is the characteristic of I.I.D assumption (*w.r.t dataset building*).
(2) In addition to the average results calculated over the entire BP range, the prediction results on each BP interval should also be reported, especially those areas of hypotension and hypertension (*w.r.t evaluation metrics*).
(3) In additional to BP, the test set should be diverse with respect to age, height, sex, etc. Especially, for video based methods, the test set should cover a large range of BP and lighting conditions (Steinman et al. 2021; Rong and Li 2021b) (*w.r.t test set*).

(4) Strict separation of training set and test set should be ensured in order to simulate the real environment, and the strategy of *splitting at record/subject level* is strongly recommended for experiments (*w.r.t splitting strategy*).
(5) The selection of the optimal feature subsets should be performed using only training set during feature selection/reduction process (*w.r.t feature selection*).
(6) Only the training set should be used to solve the normalizer during the data normalization process (*w.r.t normalization*).
(7) In addition to the conventional evaluation standards (BHS, AAMI, etc.) and analyzing tools (scatter plot, Bland-Altman plot, etc.), the analyzing tools proposed by Mukkamala et al. (2021) is strongly recommended for further evaluation (*w.r.t evaluation standards*).
(8) External evaluation is recommended to further evaluate the performance of model on other databases (*w.r.t external evaluation*).

7.4 Conclusions and future work

Future work There are many open issues worth studying.

(1) The imbalance regression is a key issue in BP estimation. However, as mentioned in Sect. 5.1.1, it has not received enough attention at present. Recall that our objective is to train an unbiased BP estimator from the severely skewed training data, similar to class-imbalanced issue, there should have data-level methods (such as under-sampling, over-sampling, etc.) and algorithm-level methods (cost-sensitive learning, etc.), which is left for further exploration.
(2) It makes sense to solve the problems of data privacy and model deployment through some cutting-edge ML technologies, such as federated learning and On-device ML, etc., both from the academic and practical point of view. Although there have been plenty of studies utilizing ML and DL technologies for BP estimation, and some promising results have been made, they fall within the scope of proof of concept. There are at least two issues that need to be considered before it can be put into practical application. Firstly, compared to experimental environment, the limited memory and weak computing capability of local computing node propose new requirements to time delay, model size and model complexity. Secondly, user privacy protection is becoming more and more important. How to collect data for model training & update without violating user's privacy is a problem that can not be ignored.
(3) It is time and necessary to explore relevant evaluation standards as well as clinical approval criterion that suitable to cuffless BP estimator, especially to those estimators based on ML and DL methods. By examining the whole process of establishing BP prediction pipeline, we revealed potential factors leading to the unreliability of results related to traditional assessment criteria such as the AAMI and the BHS standards. Besides, Mukkamala et al. (2021) has revealed the potentially misleading facts of some reported conclusions by presentating the limitations of widely-used, conventional BP

evaluation standards such as AAMI, etc., and related analyzing tools such as Regression plot and Bland-Altman plot.

(4) Long-term BP prediction (Su et al. 2018) is a challenging while meaningful direction in BP research community. Due to the time varying nature of physiological signal originating from the complex regulation mechanism of human body and the effect of abnormal event/reaction, individual's BP pattern may change over time. Therefore, the mainstream methods that under static environment may no longer be suitable. Sequence prediction and online/incremental learning scenarios may be promising solutions to long-term BP prediction.

(5) All kinds of ML algorithms and a variety of dazzling neural networks have been developed for BP prediction, and some promising results have been made. However, these approaches, especially DL methods, act as a black-box, and we still lack a clear understanding of the nature of the relationship between the input (signal) and BP. Besides, there is no further clinical validation and interpretation for the predictions. We believe that solving this problem requires the cooperation of experts from different disciplines, including machine learning, artificial intelligence, medicine and physiology. In fact, exploration of the physical principles and related models of hemodynamics help to find the most relevant factor responsible for BP change, which in further helps to identify suitable inputs and even to guide the design of informative features account for BP estimation. For example, PIR (Ding and Zhang 2015), which reflects the arterial diameter change, was initially proposed to overcome the limitation in classical PTT methods that arterial geometries keep unchanged during cardiac cycle. Womersley number (Thambiraj et al. 2019) was proposed to model the viscous flow properties of blood. Both PIR and Womersley number have been validated as significant factors in improving BP estimation accuracy (Thambiraj et al. 2019, 2020). In addition, how to combine the explicit analytical model with the deep learning model so that the latter can make more physically consistent predictions is a promising exploratory point.

Conclusions In this survey, we made a systematic review of current progress in the application of ML and DL for BP estimation, from a total of four perspectives. Especially, the content covers the whole BP prediction pipeline including dataset, signal denoising, data cleaning, feature engineering, feature selection, training algorithms, hyper-parameter optimization, evaluation procedures and evaluation metrics, etc. In addition, we discussed several critical issues and summarized several practical techniques emerging in the BP estimation community. Moreover, we introduced the potential application of several advanced ML topics in BP estimation.

Based on the significant difference about the BP prediction results reported in a large count of studies, we analyzed the factors that led to the unreliability of the results reported in some literatures by checking the whole BP prediction pipeline, from the perspective of an ML researcher. Finally, we proposed an overall proposal for an objective evaluation of different prediction methods. We accept that the proposal is not complete

and should be viewed as suggestions for further discussion by the community. Besides, how to describe the results accurately for further objective comparison is a problem worth of attention. It is certain that the previous practice of focusing mainly on the results while ignoring the data itself (data complexity, data scale in terms of the number of subjects/records/samples included, BP range, BP distribution, and the differences between training set and test set) is obviously problematic for data-driven methods. On this point, our view is consistent with the latest published review article (Liang et al. 2022) on trustworthy artificial intelligence. As shown in Table 22, we made a comprehensive comparison of relevant studies based on thirteen indicators, which provides a helpful reference for solving this problem.

In conclusion, we hope this survey can provide researchers with a systematic, comprehensive understanding of this field, including the latest advances as well as some common issues and newly-emerging techniques, and shed some light on the future directions. Meanwhile, we believe that *training a general BP predictor with genuine strong generalization ability is still challenging*, instead of the overly optimistic conclusions claimed in some literatures. In fact, the latest evaluation of smartphone-based BP estimator in a large clinical settings indicates that no commercialization has been made yet (Dörr et al. 2021). We appeal an objective view and deeper thinking on the reported results in a more systematic way.

Appendix 1: Summary of representative surveys of blood pressure prediction

We summarized several representative surveys on blood pressure measurement in Table 19.

Table 19 Representative surveys on blood pressure measurement

Title	Publication	Description
Evaluation of the accuracy of cuffless blood pressure measurement devices: Challenges and proposals Mukkamala et al. (2021)	Hypertension, 2021	Disclose the limitations of traditional evaluation standards for the validation of cuffless BP measurement methods
Accuracy of cuff-measured blood pressure: systematic reviews and meta-analyses Picone et al. (2017)	J. Am. Coll. Cardiol., 2017	Review of cuff-based BP measurements
Blood pressure measurement: clinic, home, ambulatory, and beyond Drawz et al. (2012)	Am. J. Kidney. Dis., 2012	Focus on the clinical utility of each type of measurement
Cuffless single-site photoplethysmography for blood pressure monitoring Hosanee et al. (2020)	J. Clin. Med., 2020	Review of PPG-based traditional ML methods for BP prediction
The machine learnings leading the cuffless PPG blood pressure sensors into the next stage Chao et al. (2021)	IEEE Sens. J., 2021	Review of traditional feature-based ML methods and DL methods
A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure El-Hajj and Kyriacou (2020b)	BSPC, 2020	Review of PTT/PWV/PAT/PWA methods and traditional feature- based ML methods, not covering DL methods
A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data Martinez-Ríos et al. (2021)	BSPC, 2021	Review of classification-based and regression-based methods for BP prediction
Oscillometric blood pressure estimation: Past, present, and future Forouzanfar et al. (2015)	RBME, 2015	Review of Oscillometric BP estimation
Cuffless blood pressure monitors: Principles, standards and approval for medical use Tamura (2021)	IEICE T Commun., 2021	Described the principles of cuffless BP monitors, discussed the current situation about BP monitor standards and approval for medical use
Smartphones and video cameras: Future methods for blood pressure measurement Steinman et al. (2021)	Front. Digit. Health, 2021	Review of video-based BP measurement
A survey: From shallow to deep machine learning approaches for blood pressure estimation using biosensors Maqsood et al. (2022)	Expert Syst. Appl., 2022	Review of traditional feature-based ML methods and DL methods

Appendix 2: Graphical illustration of the final processed dataset

Figure 26 presents the total BP distribution of the final processed dataset. Figure 27 illustrates the individual BP dynamics.

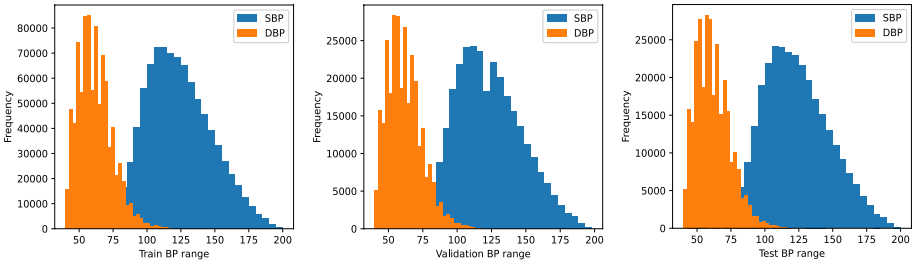

Fig. 26 Blood pressure distribution of training, validation and test set using sample level splitting strategy-③

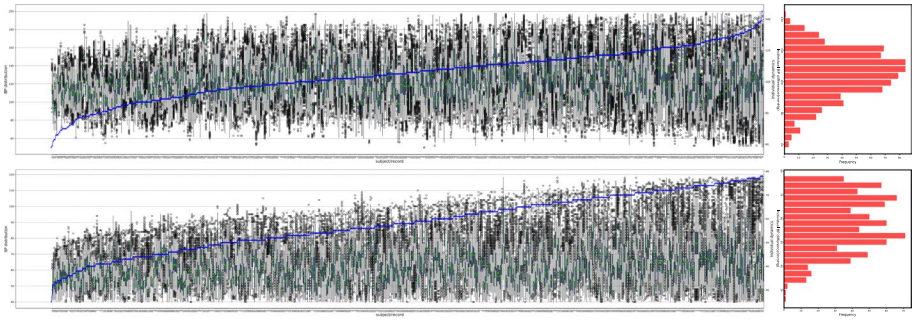

Fig. 27 Individual BP dynamics. **a** individual SBP dynamics. SBP distribution of each record is illustrated with 'Boxplot', and records are sorted in the ascending order of 'max(SBP)–min (SBP)'. The distribution of SBP dynamics of all individuals illustrated based on "Histplot" is shown in the right figure; **b** individual DBP dynamics

Appendix 3: Prediction performance of the model using different splitting strategies on different BP intervals

Figures 28 and 29 illustrates the prediction performance of the model using different splitting strategies on different BP intervals.

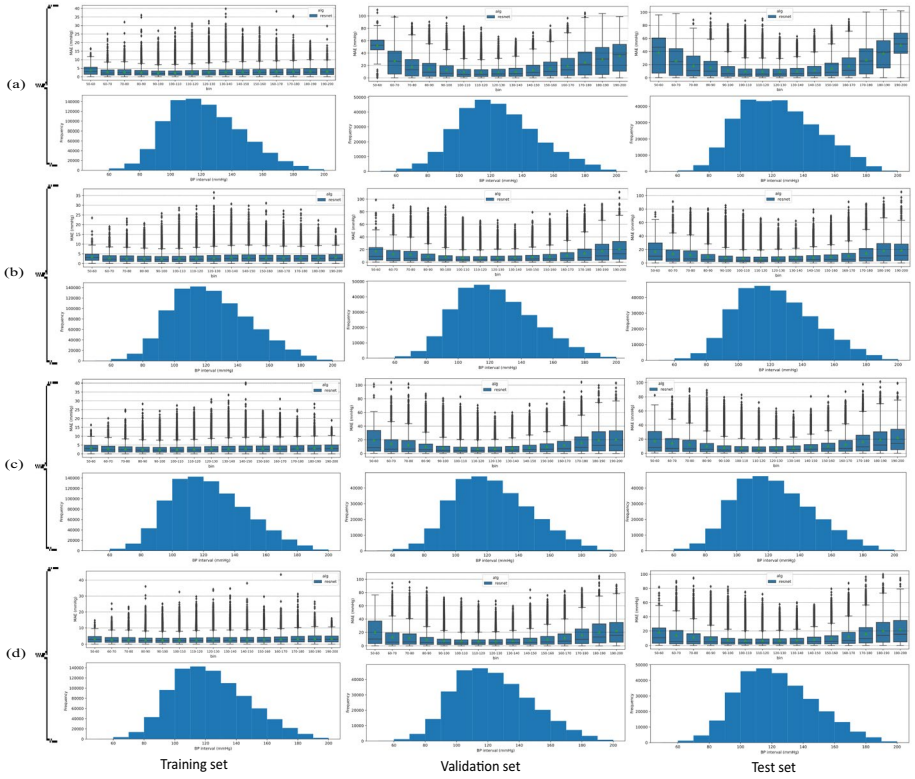

Fig. 28 Performance of ResNet model for SBP prediction on different BP intervals, based on different splitting strategies. **a** based on splitting strategy-①; **b** based on splitting strategy-③; **c** based on splitting strategy-④; **d** based on splitting strategy-⑤

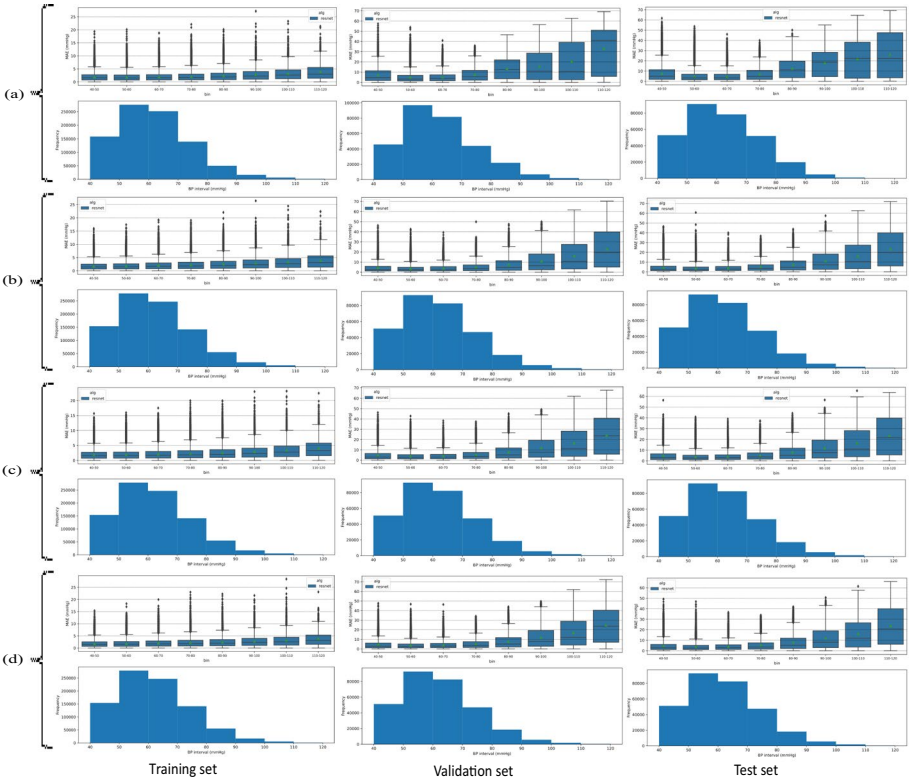

Fig. 29 Performance of ResNet model for DBP prediction on different BP intervals, based on different splitting strategies. **a** based on splitting strategy-①; **b** based on splitting strategy-③; **c** based on splitting strategy-④; **d** based on splitting strategy-⑤

Appendix 4: Open-source implementations

Table 20 summarizes the open-source implementation of all the papers we found for BP estimation.

Table 20 Summary of open-source implementations for BP prediction

Studies	Description	Link
Ibtehaz and Rahman (2020)	U-Net-based model for reconstructing ABP waveform using PPG signal	Data &code: https://github.com/nibtehaz/PPG2ABP
Aguirre et al. (2021)	Seq2Seq model with attention mechanism for reconstructing ABP waveform using PPG signal	Processed dataset: https://doi.org/10.5281/zenodo.4598938, Code: https://github.com/AguirreNicolas/ PPG2IABP
Huang et al. (2022)	MLP-Mixer based model for predicting BP using PPG and ECG signals	Data &code: https://github.com/ marshb/MLP-BP
Ji et al. (2022)	Dendritic neural model for predicting BP using PPG and ECG signals	Code: http://www.dnm.net.cn/index.html
Zhang et al. (2020b)	Domain-adversarial model for predicting BP using bioimpedance sensors	Code: https://github.com/stmilab/cufflessbp_dann
Slapničar et al. (2019)	Spectro-temporal neural network model for predicting BP using PPG signal	Code: https://github.com/gasper321/bp-estimation-mimic3
Schrumpf et al. (2021a, b)	Convolution-based model for predicting BP using PPG and rPPG signals	Code: https://github.com/Fabian-Sc85/non-invasive-bp-estimation-using-deep- learning, Processed dataset: https://zenodo.org/record/5590603
Su et al. (2018)	Deep RNN model for predicting long-term BP using PPG and ECG signals	Code: https://github.com/psu1/DeepRNN
Esmaelpoor et al. (2020)	Two-stage hybrid model for predicting BP using PPG signal	Code: https://github.com/jesmaelpoor/Blood-pressure-estimation-Deep-multistage-model
Gao et al. (2016)	SVM model for predicting BP using PPG signal	Code: https://github.com/thmedialab/DataDrivenBP
Brophy et al. (2021)	CycleGAN based model with federated learning for predicting ABP waveform using PPG signal	Code: https://github.com/Brophy-E/T2TGAN
Lo et al. (2017)	LSTM based model for predicting BP using PPG and ECG signals	Code: https://github.com/ploymel/estimateBP
Hill et al. (2021)	V-Net based model for predicting ABP waveform using PPG and ECG signals	Code: https://github.com/brianhill11/ABPImputation
Thambiraj et al. (2020)	Random forest model with genetic algorithm for predicting BP using PPG and ECG signals	data &code: https://github.com/jeya-maria-jose/Cuff_less_BP_Prediction

Appendix 5: The effect of the size of dataset on final performance

To quantitatively evaluate the effect of the size of dataset used on the experimental results, we created five versions of datasets namely v_0, v_1, v_2, v_3, and v_4, where the number of records included in v_α is α times of v_1. The corresponding statistical information of these datasets is summarized in Table 17. All experiments were performed under the same experimental settings, except that Batchsize is set to 64, 128, 256, 384, 512 for the fives versions of datasets, respectively. Inspired by Goyal et al. (2017), the corresponding initial learning rate is set to $lr = lr_0 \cdot \text{BatchSize}/128$, lr_0 equals 0.001. The maximum number of epochs is set to 50, and model is trained using Adam optimizer. Experimental code was implemented using Python 3.8 with TensorFlow 2.4.0 framework, and all the experiments were performed on Ubuntu 20.04 server equipped with two RTX 3090 GPUs.

Table 21 presents the numerical results. It is observed that the test performance drops gradually with the increase of the size (in terms of the number of records) of the dataset used, although the overall BP range and BP distribution of different versions of data sets are basically similar (refer Table 17). We attribute this to individual differences. The more records contained in the data set, the more complex and diverse individual physiological dynamics, which undoubtedly increases the difficulty of training more general models. The mean rank of each model over five experiments is calculated and the resulting p-value is 6.98e-3 (< 0.05), 2.85e-3 (< 0.05) for SBP and DBP prediction, respectively. Therefore, the null hypothesis is rejected at $\alpha = 0.05$, i.e the size of dataset has a significant effect on the trained model's performance for both SBP and DBP prediction.

Table 21 Comparison of the test performance of ResNet models on MIMIC III dataset of different size, based on record level splitting strategy-①

Method	Dataset	Task	Metrics					
			MAE	MAPE	MSE	ME	STD	R^2
ResNet	**v0**:375 records, 75e4 samples	SBP	11.76	0.100	315.164	− 0.430	17.741	0.447
		DBP	6.493	0.107	92.356	0.050	9.606	0.357
	v1:750 records, 15e5 samples	SBP	12.288	0.103	330.257	− 0.280	18.170	0.456
		DBP	6.571	0.107	91.823	0.070	9.579	0.396
	v2:1500 records, 3e6 samples	SBP	12.783	0.108	341.866	0.161	18.484	0.421
		DBP	7.080	0.115	100.765	0.575	10.037	0.357
	v3:2250 records, 4.5e6 samples	SBP	14.810	0.126	397.920	− 0.183	19.946	0.331
		DBP	8.068	0.132	116.863	− 0.236	10.808	0.258
	v4:3000 records, 6e6 samples	SBP	14.696	0.125	399.584	− 0.099	19.986	0.332
		DBP	7.955	0.130	114.348	− 0.149	10.692	0.247

Appendix 6: Comprehensive comparison of studies related to BP prediction

A comprehensive comparison of related studies for BP prediction is presented in Table 22.

Table 22 Comparison of several representative studies for BP prediction

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Slapničar et al. (2019)	Sensors,2019	MIMIC-III	510 subjects, about 500,000 samples	ICU patients	PPG, ABP	CNN+GRU	MTL	Raw signal	No	*LOSO*	No	No	15.41	12.38	–
												Yes, 20%	9.43	6.88	–
Kachuee et al. (2016)	TBME,2016	MIMIC-II	3663 records, 3663 samples	ICU patients	PPG, ECG, ABP	AdaBoost	STL	Hand-crafted features	No	*rl*, 9:1 ①	No	No	11.17	5.35	5.92
											No	Yes, single sample	8.21	4.31	–
Bose and Kandaswamy (2018)	ICACCS,2017	MIMIC-II (UCI-BP)	105 records, 8380 samples	ICU patients	PPG, ABP	Dictionary learning-RF	STL	Raw signal with single cycle	No	*rl*, 73:32 ①	No	No	17.08	10.77	–
												Yes	5.04	2.99	–
Miao et al. (2019)	JBHI,2019	Privately collected	85 subject, 2720 samples	Outpatients	ECG, PPW, cuff-based BP	MLR	STL	Hand-crafted features	No	Individual test	No	–	6.13	5.52	6.03
Simjanoska et al. (2020)	Inf. Fusion.,2020	4 data sources	51 subjects, 3129 samples	Healthy, patients	ECG, ABP	Ensemble (RF)	STL	Hand-crafted features	No	*rl*, (0.75:0.25) 0.85:0.15 ①	No	No	16.60	9.24	9.80
										sl,(0.75:0.25) 0.85:0.15 ③		–	7.93	6.41	5.72
Fan et al. (2019)	TII,2019	MIMIC II	Unknown	ICU patients	ECG, ABP	LSTM-FFNN, PSO	MTL	Raw signal	No	*rl*, 8:1:1 ①	No	No	7.16	3.89	4.24

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Fan et al. (2021)	Sensors,2021	MIMIC II	Unknown	ICU patients	ECG, ABP	LSTM-FFNN	MTL	Raw signal	No	*rl*, 8:1:1 ①	No	No	7.69	4.36	4.76
Mousavi et al. (2019b)	BSPC,2019	MIMIC II (UCI-BP)	1323 records, 3969 samples	ICU patients	PPG, ABP	ML(AdaBoost)	STL	Whole-based features	No	*sl*, 9:1 ③	No	–	3.97	2.43	2.61
Su et al. (2018)[c]	BHI,2018	Privately collected	84 subjects,	Healthy	PPG, ECG, ABP	Neural network	MTL	Hand-crafted features	No	*sl*	Yes	–	3.73*	2.43*	-
Haddad et al. (2021)	JBHI,2020	MIMIC I	28 records,	ICU patients	PPG, ABP	MLR	STL	Hand-crafted features	No	*rl*	No	Yes	6.10	4.65	4.32
Leitner et al. (2021)	JBHI,2021	MIMIC-III	100 subject records, over 10 h data	ICU patients	PPG, ABP	CNN-RNN	MTL	Raw signal	No	*rl*	No	No	16.3	8.46	–
												Yes	3.52	2.20	–
										Individual test		Yes	4.59	2.72	–
Baek et al. (2019)	IEEE Access,2019	MIMIC-II (UCI-BP)	1912 subject records	ICU patients	PPG, ECG, ABP	CNN	MTL	Raw signal	No	*rl*	No	No	9.30	5.12	–
												Yes	5.32	3.38	–
Zhang et al. (2020b)[c]	MLHC,2020	Privately collected	11 subject records	Unknown	PPG, ABP	domain adversarial Neural network	MTL	Raw signal	No	*rl*	Yes	No	7.46	4.68	–
												Yes, 5 min	6.79	4.48	–

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Schrumpf et al. (2021a)	Sensors,2021	MIMIC-III	4375 subjects, 1.5e6 samples	ICU patients	PPG, ABP	CNN	MTL	Raw signal	No	*sl*	No	–	7.7	4.4	–
			750 subjects, 1.5e6 samples							*rl*		No	16.4	8.5	–
			20 subjects							*rl*		Yes, 20% random	13.0	6.3	–
			20 subjects							*rl*		Yes, 20% first	12.3	5.8	–
Panwar et al. (2020)	IEEE Sensors,2020	MIMIC II (UCI-BP)	1557 records	ICU patients	PPG, ABP	CNN-LSTM	MTL	Raw signal	No	Unknown	Yes	–	3.97	2.30	–
Song et al. (2019)	IEEE TIM, 2019	privately collected	110 subjects	Unknown	PPG, ECG, ABP	BNN (FFNN)	STL	Hand-crafted features	Yes	*rl*	Yes	–	4.8	4.8	–
Esmaili et al. (2017)	IEEE TIM, 2017	privately collected	32 subjects	Healthy	PPG, PCG, OMRON BP	PTT	STL	Hand-crafted features	No	Individual test	No	–	6.22	3.97	–
					PPG, ECG, OMRON BP	PAT							4.471	4.44	–
Hasanzadeh et al. (2019)	IEEE Sensors, 2019	MIMIC II (UCI-BP)	Unknown	ICU patients	PPG, ABP	AdaBoost	STL	Hand-crafted features	No	*rl*	Yes	–	8.22	4.17	4.58
Mousavi et al. (2019b)	BSPC, 2019	MIMIC II (UCI-BP)	1323 records	ICU patients	PPG, ABP	AdaBoost	STL	Whole-based features	No	*sl*	No	–	3.97	2.43	2.61

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Miao et al. (2020)	Artif. Intell. Med., 2020	MIMIC III	1711 subjects, 897743 heartbeats	ICU patients	ECG, ABP	CNN + LSTM	MTL	Raw signal	No	*rl*	No	Yes	7.10	4.61	4.66
Rong and Li (2021a)	BSPC, 2021	MIMIC II (UCI)	11546 samples	ICU patients	ECG, ABP	CNN + LSTM	MTL	Raw signal	No	Unknown	No	Yes	5.59	3.36	–
Qiu et al. (2021)	BSPC, 2021	MIMIC II (UCI-BP)	1216 records	ICU patients	PPG, ECG, ABP	CNN (2D)-FFNN	MTL	Raw signal	No	*sl*	No	–	3.70	2.81	–
			40 records										1.37	0.93	–
Lin et al. (2021a)	BSPC, 2021	MIMIC II/III	109 subjects	ICU patients	PPG, ECG, ABP	MLR	STL	Hand-crafted features	No	Individual test	No	–	4.59	2.47	–
El-Hajj and Kyriacou (2021a)	BSPC, 2021	MIMIC II	500 records	ICU patients	PPG, ABP	Deep LSTM-attention	STL	Hand-crafted features	No	Unknown	No	–	2.58	1.26	–
Thambiraj et al. (2020)	BSPC, 2020	MIMIC II (UCI-BP)	Unknown	ICU patients	PPG, ECG, ABP	RF	STL	Hand-crafted features	No	*rl*	No	No	9.54	5.48	3.27
Attarpour et al. (2019)	BSPC, 2019	MIMIC II (UCI-BP)	111 individuals, 111 samples	Healthy	PPGs, OMRON-based BP	MLP	STL	Hand-crafted features	No	LOSO	No	No	5.59	4.45	–
Dagamseh et al. (2021)	Biomed. Opt. Express, 2021	Privately collected	448 subjects, 448 samples	Normal+ abnormal	PPGs, mercury-based BP	SVM	STL	Hand-crafted features	No	LOSO	No	No	4.69	3.39	–

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Baker et al. (2021)	Comput. Meth. Prog. Bio., 2021	MIMIC III	6972 subjects, 200000 samples	ICU patients	PPG, ECG, ABP	CNN-LSTM	STL	Raw signal	No	Unknown	No	–	4.41	2.91	2.77
Fong et al. (2019)	Comput. Biol. Med., 2019	Privately collected	40 subjects	Healthy	PPGs, OMRON-based BP	Ensemble (SVR)	STL	Hand-crafted features	No	LOSO	No	No	7.29	5.01	–
Esmaelpoor et al. (2020)	Comput. Biol. Med., 2020	MIMIC II	200 subjects, 51884 heartbeats	ICU patients	PPG, ABP	Two stage, CNN-LSTM	MTL	Raw signals	No	*sl*	No	–	3.97	2.10	–
Schlesinger et al. (2020)	ICASSP, 2020	MIMIC II	304 subjects, 106074 samples	ICU patients	PPG, ABP	STFT+Siamese(2D CNN)	STL	Raw signal	No	*sl*,6:2:2 ③	No	Yes	5.95	3.41	–
Schlesinger et al. (2020)	Crit. Care Explor., 2020	MIMIC II	329 subjects, 136459 samples	ICU patients	PPG, ABP	STFT+Siamese(2D CNN)	STL	Raw signal	No	*rl*,6:2:2 ①	No	Yes	9.71	4.95	–
Baek et al. (2020)	EMBC, 2020	Privately collected	26 subjects	Unknown	MwPPG, auscultatory BP	CNN	MTL	Raw signal	No	*sl*, 6:1:2 ④	No	–	5.28	4.95	6.5
Tazarv and Levorato (2021)	EMBC, 2021	MIMIC II	20 subjects	ICU patients	PPG, ABP	CNN-LSTM	Unknown	Raw signal	No	*sl*	Yes	–	3.70	2.02	–
		QUVD	490 min	Surgical patients									3.91	1.99	–

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Liu et al. (2021)	IEEE Access, 2021	Privately collected	62 subjects, 101270 heartbeats	Normal + abnormal	PPG, ECG, mercury -based BP	GPR	STL	Hand-crafted features	No	*sl*	No	–	4.8	3.4	–
Zhang et al. (2021c)	IEEE Access, 2021	Privately collected	20 subjects	Healthy	PPG, ECG, BCG, cuff BP	LR	STL	Hand-crafted features	No	Unknown	No	?	6.84	5.46	–
Chen et al. (2021)	IEEE Access, 2021	Privately collected	1060 groups	Unknown	PPG, ECG, BCG, cuff BP	RF	STL	Hand-crafted features	Yes	Unknown	No	?	4.45	3.95	–
Miao et al. (2019)	JBHI, 2019	Privately collected	85 subjects, 2720 samples	Healthy	PPWs, ECG, OMRON BP	MIR	STL	Hand-crafted features	No	Individual test	No	–	6.13	4.54	4.81
Yang et al. (2020a)	JBHI, 2021	VitalDB	1376 subjects, over 240 w heartbeats	Healthy	PPG, ECG, ABP	ANN+RNN	STL	Hand-crafted features	No	*rl*, 7:3 ①	No	No	5.07	2.86	–
Liu et al. (2020a)	JBHI, 2021	Privately collected	22 subjects	Elderly individuals	MwPPG, auscultatory BP	Explicit analytical model	–	Hand-crafted features	No	Individual test	No	No	5.51	5.57	4.67
Ibrahim and Jafari (2019)	IEEE Trans. Biomed. Circ. Syst., 2019	Privately collected	10 subjects, 2848 heartbeats	Healthy	Four Bio-Z signals, Finapres NOVA-based BP	AdaBoost	STL	Hand-crafted features	No	Individual test, 8:1:1 ⑦	No	–	5.51	5.57	4.67

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Wang et al. (2018b)	J. Healthc. Eng., 2018	MIMIC	72 subjects, 58795 samples	ICU patients	PPG, ABP	FFNN (MLP)	MTL	Hand-crafted features	No	*sl*, 70:15:15 ③	No	–	4.02[g]	2.27[g]	–
										Individual test			4.41[g]	2.87[g]	–
Yang et al. (2021)	Opt. Quantum Electron., 2021	Privately collected	45 subjects, 315 records	Healthy	PPG, ECG, LBPK1-based BP	CNN	STL	Raw signals	Yes	*rl*	No	No	4.43	3.23	–
									No				5.51	4.32	–
Yamakoshi et al. (2021)	PeerJ, 2021	Privately collected	13 subjects, 299 samples	Normal + abnormal	PPG, brachial cuff-based BP	Explicit analytical model	–	Hand-crafted features	No	Individual test	No	Yes	3.6[g]	3.3[g]	4.3[g]
Zhang et al. (2021a)	Physiol. Meas.,2021	MIMIC	32 subject, 36 records	ICU patients	PPG, ECG, ABP	AdaBoost	STL	Hand-crafted features	No	*sl*	No	No	6.6	3.12	–
		VitalDB	3305 subjects, 3305 records	Healthy									11.67	6.55	–
		MIMIC + VitalDB	3337 subjects, 3341 records	ICU patients, healthy									10.03	5.42	–
Thambiraj et al. (2019)	Physiol. Meas., 2019	Privately collected	42 subjects	Healthy	PPG, ECG, OMRON BP	Explicit analytic model (PTT_w)	–	Raw signal	No	*rl*	No	No	1.40	0.75	0.55
		Privately collected	39 subjects	Diseased									1.90	0.99	0.047

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Malayeri and Khodabakhshi (2022)	Unpublished, 2022	MIMIC II (UCI-BP)	200 records	ICU patients	PPG, ABP	CNN (1D) + CNN (2D)	STL	Raw signal	No	*sl*	No	–	3.05	1.58	–
Jeong and Lim (2021)	Sci. Rep., 2021	MIMIC	48 subjects	ICU patients	PPG, ECG, ABP	CNN-LSTM	MTL	Raw signal	No	Unknown	No	–	1.2	1.0	–
Simjanoska et al. (2018)	Sensors,2018	Privately collected	51 subjects from four sources, 3129 samples	Healthy (33), diseased (18)	ECG, ABP	RF	STL	Hand-crafted features	Yes	*rl*	No	No	8.64	18.20	13.52
												Yes	7.72	9.45	8.13
Shao et al. (2020)	Sensors, 2020	Privately collected	21 subjects, 100000 heartbeats	Healthy	PPG, ECG, cuff BP	Explicit analytical model (PTT)	–	Hand-crafted features	No	Individual test	No	Yes, mPTT	6.0	4.4	–
												Yes, fPTT	5.3	4.0	–
Eom et al. (2020)	Sensors,2020	Privately collected	15 subject, 26600014 samples	Normal	PPG, ECG, BCG, ABP	CNN-RNN	MTL	Raw signal	No	*sl*, 7:1:2	No	–	4.06	3.33	–
										LOSO		No	9.70	5.79	–
Chowdhury et al. (2020)	Sensors, 2020	PPGBP	126 subjects, 222 samples	Outpatients	PPG, ABP	GPR	STL	Hand-crafted features	Yes	*sl*	No	–	3.02	1.74	–
Hsu et al. (2020)	Sensors, 2020	MIMIC II (UCI-BP)	9000 records, 2176188 heartbeats	ICU patients	PPG, ABP	FFNN	MTL	Hand-crafted features	No	*sl*,7:1:2	No	–	3.21	2.23	–

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Li et al. (2020a)	Sensors, 2020	MIMIC II (UCI-BP)	50 records, 6852 heartbeats	ICU patients	PPG, ECG, ABP	deep LSTM	MTL	Hand-crafted features	No	*sl*, 4:1	No	–	0.7357	0.5587	–
			3e3 records, 678202 heartbeats										6.726	2.516	–
Wang and Lin (2020)	Sensors, 2020	Orivately collected	30 subjects	Healthy	PPW, oscillometric BP	Explicit analytical model	–	Pressure change	No	Individual test	No	Yes	1.52	1.83	–
Lee et al. (2021)	Sensors,2021	Privately collected	18 subjects	Unknown	PPG, ECG, BCG, ABP	LSTM	MTL	Hand-crafted features	Yes	*rl*	No	No	10.01	5.64	–
												Yes, 20%	2.56	2.05	–
										Individual test		Yes	2.62	2.03	–
Hassani and Foruzan (2019)	Signal Image Video P., 2019	MIMIC II	120 subjects	ICU patients	PPG,, ABP	MLP-SVR	STL	Hand-crafted features	No	Unknown	No	?	1.21	0.80	–
Fati et al. (2021)	Symmetry, 2021	MIMIC	120 subjects, 64115 samples	ICU patients	PPG,, ABP	MLP-SVR	STL	Hand-crafted features	No	*sl*, 3:2 ③	No	?	6.52	4.19	–
Ertuğrul and Sezgin (2018)	Turk.J. Elec. Eng. Comp. Sci., 2018	MIMIC II (UCI-BP)	4254 records	ICU patients	PPG, ECG, ABP	ELM	Unknown	Hand-crafted features	No	*sl*	No	–	4.371	3.953	3.639
			5 records							Individual test		Yes	1.054	1.383	1.008

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Sharifi et al. (2019)	Artif. Intell. Med., 2019	MIMIC II (UCI-BP)	3663 records	ICU patients	PPG, ECG, ABP	State space reconstruction + MARS	STL	Hand-crafted features	No	*sl*, 4:1 ④	No	–	7.83	4.86	3.63
Tanveer and Hasan (2019)	BSPC, 2019	MIMIC	39 subjects	ICU patients	PPG, ECG, ABP	ANN-LSTM	MTL	Raw signals	No	*sl*, 7:1:2 ④	No	–	1.10	0.58	–
Cheng et al. (2021)	Comput. Biol. Med., 2021	MIMIC II (UCI-BP)	1620 records, 277050 samples	ICU patients	PPG, ABP	CNN (UNet)	–	Raw signal	No	*sl*, 9:1 ③	Yes	–	3.27	1.90	1.49
										rl		Yes	6.41	3.73	3.30
Sadrawi et al. (2020)	Sensors, 2020	Privately collected	18 subjects, 36516 samples	Surgical patients	PPG, ABP	CNN (UNet) + GA	–	Raw signals	No	*sl*, 85:15 ③	No	–	2.54	1.48	–
Li and He (2021)	Sensors, 2021	MIMIC II (UCI-BP)	3183 subjects, 9549 samples	ICU patients	PPG, ABP	GRNN	–	Raw signals	No	*sl*, 70:15:15 ③	No	–	3.96	2.39	–
Athaya and Choi (2021)	Sensors, 2021	MIMIC I/III	100 subject records, about 195 h duration	ICU patients	PPG, ABP	CNN (UNet)	–	Raw signals	No	Unknown, 70:15:15 ③	Yes	–	3.68	1.97	–
Aguirre et al. (2021)	Sensors, 2021	MIMIC III	1131 subjects, 10696 samples	ICU patients	PPG, ABP	Seq2Seq (GRU)	–	Hand-crafted features	Yes	*sl*	Yes	–	12.08	5.56	–
										rl			14.39	6.57	–
									No				14.55	7.01	–

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Harfiya et al. (2021)	Sensors, 2021	MIMIC II (UCI-BP)	5289 records, 250000 samples	ICU patients	PPG, ABP	AutoEncoder (LSTM)	–	Raw signal	No	Unknown	No	?	4.05	2.41	–
Mahmud et al. (2022)	Sensors, 2022	MIMIC II (UCI-BP)	5289 records, 200159 samples	ICU patients	PPG, ECG, ABP	UNet (feature learner) + MLP	–	Raw signal	No	*rl*, 3:1 ①	Yes	No	2.333	0.713	–
Qin et al. (2021)	BSPC, 2021	MIMIC II (UCI-BP)	1227 records, 39264 samples	ICU patients	PPG, ABP	Autoencoder with domain adversarial training	–	Raw signal	No	*rl*, 6:2:2 ①	Yes	No	7.945	4.114	3.834
												Yes	5.424	3.144	2.885
Li et al. (2021)	Wirel. Commun. Mob. Comput.,2021	MIMIC II (UCI-BP)	Unknown records, 50165 samples	ICU patients	PPG, ABP	CNN+Bi-LSTM	MTL	Raw signal	No	Unknown, 8(0.9:0.1):2	No	?	7.85	4.42	–
Ji et al. (2022)	IEEE Trans. Cybern., 2022	MIMIC	Unknown	ICU patients	PPG, ECG, ABP	NN (ADNR)	Unknown	Hand-crafted features	No	Unknown	No	?	2.19	1.16	1.43
Lo et al. (2017)[c]	EMBC, 2017	MIMIC II (UCI-BP)	25 records, unknown samples	ICU patients	PPG, ECG, ABP	LSTM	Unknown	Raw signals	No	Unknown, 68:12:20	No	?	2.751	1.604	–
Zheng and YuJ. (2021)	Nanomater.,2021	Privately collected	250 subjects, 250 samples	Nonhypertension individuals	Clinical data, ABP	FFNN (MLP) + BRA	STL	Clinical and lifestyle features	Yes	*rl*, 70:15:15 ②	No	No	6.28	–	–
										rl, 80:10:10 ②			5.85	–	–
										rl, 90:5:5 ②			5.16	–	–

Table 22 (continued)

Studies	Publication	Data source	#Data used[e]	Health state	Signal used	Method[d]	Task corr.[a]	Input type	Demo. Fea.	Splitting strategy[b]	Norm. target	Calibration[f]	MAE (unit:mmHg)		
													SBP	DBP	MBP
Chuang et al. (2021)	Appl. Sci., 2021	MIMIC I	45 subjects, about 4.95e5 samples	ICU patients	PPG, ECG, ABP	CNN-LSTM	Unknown	Raw signals	No	Unknown, 9:1	No	?	2.94	2.02	–

[a]According to whether correlation between different prediction tasks (SBP, DBP and MBP) is considered, related studies can be categorized into STL and MTL modes. Especially, this item is not applicable to explicit analytical model/mathematical model or those studies based on waveform reconstruction (e.g Cheng et al. (2021)).

[b]Denotes training, validation and test sets were split at the final aggregated sample level, *rl* denotes training, validation and test sets were split at the subject/record level, *LOSO* denotes leave one subject/record out, which can be seen as a special case of *rl*. Refer Table 22 for the meaning of circled number.

[c]In these works, only the results of RMSE are provided.

[d]'A−B'/'A+B' indicates that 'A' and 'B' are sequential/parallel.

[e]Note that the UCI-BP dataset published by Kachuee et al. (2015), Kachuee et al. (2016) contains a total of 12000 waveform records from 942 patients/subjects, and the subordinate relationship between records and patients is unknown. It's obviously that the concept of 'record' and 'subject' is confused in the relevant literatures (Qiu et al. 2021; Li et al. 2020a; Harfiya et al. 2021; Lo et al. 2017) when using this version of dataset. We have corrected the incorrect statement.

[f]Note that for explicit analytical model, calibration means determine a few parameters of the model using a subset of an individual's data, which can be seen as training prediction model to a certain extent. However, for ML/DL methods, calibration means fine tuning the general model trained on other individual's data using a subset of test individual's data. '–' indicates that the corresponding term is not applicable.

[g]Note the result is acquired by averaging the results from original papers

'–' denotes not applicable. *Demo. Fea.* demographical features, *MLR* multiple linear regression, *MIR* multi-instance regression, *RF* random forest, *ELM* extreme learning machine, *MARS* multi-adaptive regression spline, *GRNN* generalized regression neural network, *BRA* Bayesian Regularization Algorithm, *GA* genetic algorithm

Appendix 7: Abbreviations

All acronyms appearing in the paper are summarized in Table 23.

Table 23 Abbreviation table

Acronyms	Full name
AAMI	The Association for the Advancement of Medical Instrumentation
ABP	Arterial blood pressure
AD	After diastolic
AI	Artificial intelligence
AMPD	Automatic multi-scale-based peak detection
ANN	Artificial neural network
APG	Accelerated plethysmograph
ARIMA	Autoregressive integrated moving average model
AW	Auscultatory waveform
BCG	Ballistocardiogram
Bi-LSTM	Bidirectional long short-term memory
BH	Bramwell-Hill
BHS	British Hypertension Society
BMI	Body mass index
BO	Bayesian optimization
BRA	Bayesian regularization algorithm
BW	Baseline wandering
BNN	Boosting neural networks
BP	Blood pressure
BSD	Between systolic and diastolic
CART	Classification and regression tree
CNN	Convolutional neural networks
CO	Cardiac output
CRNN	Convolutional-recurrent neural network
CV	Computer vision
CWT	Continuous wavelet transform
DBP	Diastolic blood pressure
DCT	Discrete cosine transform
DL	Deep learning
DNM	Dendritic neural model
DNN	Deep neural network
DPI	Dynamic plosion index
DT	Decision tree
DWT	Discrete wavelet transform
ECG	Electrocardiograpshy
EDA	Electrodermal activity
ELM	Extreme learning machine
EMD	Empirical mode decomposition
EMG	Electromyogram
ERM	Empirical risk minimization
FC	Fully-connected
FFNN	Feedforward neural network

Table 23 (continued)

Acronyms	Full name
FFT	Fast Fourier transform
FRP	Fuzzy recurrent plot
GA	Genetic algorithm
GAN	Generative adversarial network
GMM-HMM	Gaussian mixture models and hidden markov model
GPR	Gaussian process regression
GRNN	Generalized regression neural network
GRU	Gate recurrent unit
HFC	High frequency components
HPO	Hyper-parameter optimization
ICG	Impedance-cardiogram
ICU	Intensive care unit
I.I.D	Independent-identical-distribution
iPTT	Image-based PTT
iPPG	Image-based PPG
JADE	Joint approximate diagonalization of eigenmatrices
K-SVD	K-singular value decomposition
LCFs	Level-crossing features
LDA	Latent dirichlet allocation
LFC	Low frequency components
LIME	Locally interpretable model-agnotic explanations
LOSO	Leave one subject out
LR	Linear regression
LSTM	Long short-term memory
LTF	Linear transfer function
ML	Machine learning
MA	Motion artifacts
MAE	Mean absolute error
MAML	Model-agnostic meta-learning
MAPE	Mean absolute percentage error
MARS	Multi-adaptive regression spline
ME	Mean error
MGD	Multivariate Gaussian distribution
MIMIC	Multiparameter Intelligent Monitoring in Intensive Care
MIR	Multi-instance regression
MK	Moens-Korteweg
MLR	Multiple linear regression
MSE	Mean square error
MTL	Multi-task learning
MWPPG	Multi wavelength PPG
MCPPG	Multi channel PPG
MFMC	Multi-filter to multi-channel
MLP	Multilayer perceptron
NARX	Nonlinear autoregressive model with exogenous input
NAS	Neural architecture search
NNOE	Neural network output-error
OWR	Online weighted resampling
OMW	Oscillometric waveform

Table 23 (continued)

Acronyms	Full name
PCG	Phonocardiogram
PS	Pre-systolic
PTT	Pulse transit time
PWV	Pulse wave velocity
PAT	Pulse arrival time
PCA	Principal components analysis
PI	Pressure index
PIR	Photoplethysmogram intensity ratio
PLS	Partial least square
PLI	Power line interference
PWA	Pulse wave analysis
PPG	Photoplethysmography
PSO	Particle swarm optimization
PZT	Piezoelectric
RF	Random forest
RFE	Recursive feature elimination
RFFS	Random forest with feature selection
RFSV	Random forest with shapley value
RNN	Recurrent neural networks
ROI	Region-of-interest
rPPG	Remote photoplethysmography
RSP	Respiratory
SBP	Systolic blood pressure
SCG	Seismocardiogram
SCSA	Semi-classical signal analysis
SHAP	Shapley values
SRM	Structural risk minimization
STD	Standard error
STL	Single-task learning
SVM	Support vector machine
SVR	Support vector regression
TOI	Transdermal optical imaging
TPOT	Tree-based pipeline optimization tool
TPR	Total peripheral vascular resistance
UCI	University of California Irvine
VPG	Velocity plethysmography
WFDB	Waveform database
WSFS	Weakly supervised feature selection

Data availability The experiments involved are based on a publicly available database MIMIC III, the script to download the dataset and the related experimental code are released on the Github repository: https://github.com/v3551G/BP-prediction-survey.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abderahman HN, Dajani HR, Bolic M, Groza VZ (2017) An integrated blood pressure measurement system for suppression of motion artifacts. Comput Methods Prog Biomed 145:1–10

Abrar S, Loo CK, Kubota N, Tahir GA (2020) A personalised blood pressure prediction system using Gaussian mixture regression and online recurrent extreme learning machine. In: 2020 international symposium on community-centric systems (CcS), pp 1–7. IEEE

Aguirre N, Grall-Maës E, Cymberknop LJ, Armentano RL (2021) Blood pressure morphology assessment from photoplethysmogram and demographic information using deep learning with attention mechanism. Sensors 21(6):2167

Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322

Ahmad S, Chen S, Soueidan K, Batkin I, Bolic M, Dajani H, Groza V (2012) Electrocardiogram-assisted blood pressure estimation. IEEE Trans Biomed Eng 59(3):608–618

Argha A, Celler BG (2019) Blood pressure estimation from time-domain features of oscillometric waveforms using long short-term memory recurrent neural networks. IEEE Trans Instrum Meas 69(6):3614–3622

Argha A, Wu J, Su SW, Celler BG (2019) Blood pressure estimation from beat-by-beat time-domain features of oscillometric waveforms using deep-neural-network classification models. IEEE Access 7:113427–113439

Argha A, Celler BG, Lovell NH (2020) A novel automated blood pressure estimation algorithm using sequences of Korotkoff sounds. IEEE J Biomed Health Inform 25(4):1257–1264

Arsene CT, Hankins R, Yin H (2019) Deep learning models for denoising ECG signals. In: 2019 27th European signal processing conference (EUSIPCO), pp 1–5. IEEE

Athaya T, Choi S (2021) An estimation method of continuous non-invasive arterial blood pressure waveform using photoplethysmography: a U-Net architecture-based approach. Sensors 21(5):1867

Atomi K, Kawanaka H, Bhuiyan M, Oguri K et al (2017) Cuffless blood pressure estimation based on data-oriented continuous health monitoring system. Comput Math Method Med

Attarpour A, Mahnam A, Aminitabar A, Samani H (2019) Cuff-less continuous measurement of blood pressure using wrist and fingertip photo-plethysmograms: evaluation and feature analysis. Biomed Signal Process Control 49:212–220

Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM et al (2019) Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med 25(1):70–74

Baek S, Jang J, Yoon S (2019) End-to-end blood pressure prediction via fully convolutional networks. IEEE Access 7:185458–185468

Baek S, Jang J, Cho S-H, Choi JM, Yoon S (2020) Blood pressure prediction by a smartphone sensor using fully convolutional networks. In: 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC), pp 188–191. IEEE

Baker S, Xiang W, Atkinson I (2021) A hybrid neural network for continuous and non-invasive estimation of blood pressure from raw electrocardiogram and photoplethysmogram waveforms. Comput Methods Prog Biol 207:106191

Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. J Mach Learn Res 13(2)

Bishop CM, Nasrabadi NM (2006) Pattern recognition and machine learning, vol 4. Springer, Berlin

Block RC, Yavarimanesh M, Natarajan K, Carek A, Mousavi A, Chandrasekhar A, Kim C-S, Zhu J, Schifitto G, Mestha LK et al (2020) Conventional pulse transit times as markers of blood pressure changes in humans. Sci Rep 10(1):1–9

Bose SSN, Kandaswamy A (2017) Sparse representation of photoplethysmogram using K-SVD for cuffless estimation of arterial blood pressure. In: 2017 4th international conference on advanced computing and communication systems (ICACCS), pp 1–5 . IEEE

Bose SSN, Kandaswamy A (2018) Sparse characterization of PPG based on K-SVD for beat-to-beat blood pressure prediction. Biomed Res 29:835–843. https://doi.org/10.4066/biomedicalresearch.29-17-1438

Bromley J, Bentz JW, Bottou L, Guyon I, LeCun Y, Moore C, Säckinger E, Shah R (1993) Signature verification using a Siamese time delay neural network. Int J Pattern Recogn Artif Intell 7(04):669–688

Brophy E, De Vos M, Boylan G, Ward T (2021) Estimation of continuous blood pressure from PPG via a federated learning approach. Sensors 21(18):6311

Buchanan S, Orris P, Karliner J (2011) Alternatives to the mercury sphygmomanometer. J Public Health Policy 32(1):107–120

Butt MM, Akram U, Khan SA (2015) Denoising practices for electrocardiographic (ECG) signals: a survey. In: 2015 international conference on computer, communications, and control technology (I4CT), pp 264–268. IEEE

Celler BG, Le PN, Argha A, Ambikairajah E (2019a) Blood pressure estimation using time domain features of auscultatory waveforms and GMM-HMM classification approach. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 208–211. IEEE

Celler BG, Le PN, Argha A, Ambikairajah E (2019b) GMM-HMM-based blood pressure estimation using time-domain features. IEEE Trans Instrum Meas 69(6):3631–3641

Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40(1):16–28

Chandrasekaran V, Dantu R, Jonnada S, Thiyagaraja S, Subbu KP (2012) Cuffless differential blood pressure estimation using smart phones. IEEE Trans Biomed Eng 60(4):1080–1089

Chandrasekhar A, Yavarimanesh M, Natarajan K, Hahn J-O, Mukkamala R (2020) Ppg sensor contact pressure should be taken into account for cuff-less blood pressure measurement. IEEE Trans Biomed Eng 67(11):3134–3140

Chao PC-P, Wu C-C, Nguyen DH, Nguyen B-S, Huang P-C, Le V-H (2021) The machine learnings leading the cuffless PPG blood pressure sensors into the next stage. IEEE Sens J 21(11):12498–12510

Chen W, Kobayashi T, Ichikawa S, Takeuchi Y, Togawa T (2000) Continuous estimation of systolic pressure using the pulse arrival time and intermittent calibration. Med Biol Eng Comput 38:569–574. https://doi.org/10.1007/BF02345755

Chen S, Ji Z, Wu H, Xu Y (2019) A non-invasive continuous blood pressure estimation approach based on machine learning. Sensors 19(11):2585

Chen X, Yu S, Zhang Y, Chu F, Sun B (2021) Machine learning method for continuous noninvasive blood pressure detection based on random forest. IEEE Access 9:34112–34118

Chen J-W, Huang H-K, Fang Y-T, Lin Y-T, Li S-Z, Chen B-W, Lo Y-C, Chen P-C, Wang C-F, Chen Y-Y (2022) A data-driven model with feedback calibration embedded blood pressure estimator using reflective photoplethysmography. Sensors 22(5):1873

Cheng J, Xu Y, Song R, Liu Y, Li C, Chen X (2021) Prediction of arterial blood pressure waveforms from photoplethysmogram signals via fully convolutional neural networks. Comput Biol Med 138:104877

Chiang P-H, Dey S (2018) Personalized effect of health behavior on blood pressure: Machine learning based prediction and recommendation. In: 2018 IEEE 20th international conference on e-health networking, applications and services (IEEE Healthcom), pp 1–6. IEEE

Chiang P-H, Dey S (2019) Offline and online learning techniques for personalized blood pressure prediction and health behavior recommendations. IEEE Access 7:130854–130864

Chiang P-H, Wong M, Dey S (2021) Using wearables and machine learning to enable personalized lifestyle recommendations to improve blood pressure. IEEE J Transl Eng Health Med 9:1–13

Chopra S, Hadsell R, LeCun Y (2005) Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR), vol 1, pp 539–546. IEEE

Chowdhury MH, Shuzan MNI, Chowdhury ME, Mahbub ZB, Uddin MM, Khandakar A, Reaz MBI (2020) Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques. Sensors 20(11):3127

Chuang C-C, Lee C-C, Yeng C-H, So E-C, Chen Y-J (2021) Attention mechanism-based convolutional long short-term memory neural networks to electrocardiogram-based blood pressure estimation. Appl Sci 11(24):12019

Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555

Cornelissen VA, Smart NA (2013) Exercise training for blood pressure: a systematic review and meta-analysis. J Am Heart Assoc 2(1):004473

Dagamseh A, Qananwah Q, Al Quran H, Ibrahim KS (2021) Towards a portable-noninvasive blood pressure monitoring system utilizing the photoplethysmogram signal. Biomed Opt Express 12(12):7732–7751

Das N, Chakraborty M (2017) Performance analysis of FIR and IIR filters for ECG signal denoising based on SNR. In: 2017 third international conference on research in computational intelligence and communication networks (ICRCICN), pp 90–97. IEEE

Das M, Choudhary T, Sharma L, Bhuyan M (2020) Accelerometric method for cuffless continuous blood pressure measurement. arXiv preprint arXiv:2008.07899

Dash A, Ghosh N, Patra A, Choudhury AD (2020) Estimation of arterial blood pressure waveform from photoplethysmogram signal using linear transfer function approach. In: 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC), pp 2691–2694. IEEE

Dastjerdi AE, Kachuee M, Shabany M (2017) Non-invasive blood pressure estimation using phonocardiogram. In: 2017 IEEE international symposium on circuits and systems (ISCAS), pp 1–4. IEEE

Datta S, Banerjee R, Choudhury AD, Sinha A, Pal A (2016) Blood pressure estimation from photoplethysmogram using latent parameters. In: 2016 IEEE international conference on communications (ICC), pp 1–7 . IEEE

Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

Dey J, Gaurav A, Tiwari VN (2018) InstaBP: cuff-less blood pressure monitoring on smartphone using single PPG sensor. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 5002–5005. IEEE

Dhar S, Guo J, Liu J, Tripathi S, Kurup U, Shah M (2021) A survey of on-device machine learning: an algorithms and learning theory perspective. ACM Trans Internet Things 2(3):1–49

Ding X-R, Zhang Y-T (2015) Photoplethysmogram intensity ratio: a potential indicator for improving the accuracy of PTT-based cuffless blood pressure estimation. In: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 398–401. IEEE

Ding X, Zhang Y-T, Liu J, Dai W-X, Tsang H (2015) Continuous cuffless blood pressure estimation using pulse transit time and photoplethysmogram intensity ratio. IEEE Trans Biomed Eng 63:964–972. https://doi.org/10.1109/TBME.2015.2480679

Ding X, Yan BP, Zhang Y-T, Liu J, Zhao N, Tsang HK (2017) Pulse transit time based continuous cuffless blood pressure estimation: a new extension and a comprehensive evaluation. Sci Rep 7(1):1–11

Ding X, Yan BP, Zhang Y-T, Liu J, Su P, Zhao N (2019) Feature exploration for knowledge-guided and data-driven approach based cuffless blood pressure measurement. arXiv preprint arXiv:1908.10245

Djeldjli D, Bousefsaf F, Maaoui C, Bereksi-Reguig F, Pruski A (2021) Remote estimation of pulse wave features related to arterial stiffness and blood pressure using a camera. Biomed Signal Process Control 64:102242

Dörr M, Weber S, Birkemeyer R, Leonardi L, Winterhalder C, Raichle CJ, Brasier N, Burkard T, Eckstein J (2021) iPhone App compared with standard blood pressure measurement-the iPARR trial. Am Heart J 233:102–108

Dragomiretskiy K, Zosso D (2013) Variational mode decomposition. IEEE Trans Signal Process 62(3):531–544

Drawz PE, Abdalla M, Rahman M (2012) Blood pressure measurement: clinic, home, ambulatory, and beyond. Am J Kidney Dis 60(3):449–462

Du P, Kibbe WA, Lin SM (2006) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics 22(17):2059–2065

Ebrahim MP, Heydari F, Wu T, Walker K, Joe K, Redoute J-M, Yuce MR (2019) Blood pressure estimation using on-body continuous wave radar and photoplethysmogram in various posture and exercise conditions. Sci Rep 9(1):1–13

Edition T, Moody GB WFDB applications guide. https://physionet.org/physiotools/wag/wag.htm

El Attaoui A, Largo S, Jilbab A, Bourouhou A (2020) Wireless medical sensor network for blood pressure monitoring based on machine learning for real-time data classification. J Ambient Intell Hum Comput 1–16

El-Hajj C, Kyriacou PA (2020a) Cuffless and continuous blood pressure estimation from PPG signals using recurrent neural networks. In: 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC), pp 4269–4272. IEEE

El-Hajj C, Kyriacou PA (2020b) A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. Biomed Signal Process Control 58:101870

El-Hajj C, Kyriacou PA (2021a) Deep learning models for cuffless blood pressure monitoring from PPG signals using attention mechanism. Biomed Signal Process Control 65:102301

El-Hajj C, Kyriacou P (2021b) Cuffless blood pressure estimation from PPG signals and its derivatives using deep learning models. Biomed Signal Process Control 70:102984

Elsken T, Metzen JH, Hutter F (2019) Neural architecture search: a survey. J Mach Learn Res 20(1):1997–2017

Eom H, Lee D, Han S, Hariyani YS, Lim Y, Sohn I, Park K, Park C (2020) End-to-end deep learning architecture for continuous blood pressure estimation using attention mechanism. Sensors 20(8):2338

Ertuğrul ÖF, Sezgin N (2018) A noninvasive time-frequency-based approach to estimate cuffless arterial blood pressure. Turk J Electr Eng Comput Sci 26(5):2260–2274

Esmaelpoor J, Moradi MH, Kadkhodamohammadi A (2020) A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals. Comput Biol Med 120:103719
Esmaelpoor J, Moradi MH, Kadkhodamohammadi A (2021a) Cuffless blood pressure estimation methods: physiological model parameters versus machine-learned features. Physiol Meas 42(3):035006
Esmaelpoor J, Sanat ZM, Moradi MH (2021b) A clinical set-up for noninvasive blood pressure monitoring using two photoplethysmograms and based on convolutional neural networks. Biomed Eng
Esmaili A, Kachuee M, Shabany M (2017) Nonlinear cuffless blood pressure estimation of healthy subjects using pulse transit time and arrival time. IEEE Trans Instrum Meas 66(12):3299–3308
Fan X, Ye Q, Yang X, Choudhury SD (2018) Robust blood pressure estimation using an RGB camera. J Amb Intell Hum Comput 1–8
Fan X, Wang H, Xu F, Zhao Y, Tsui K-L (2019) Homecare-oriented intelligent long-term monitoring of blood pressure using electrocardiogram signals. IEEE Trans Ind Inform 16(11):7150–7158
Fan X, Wang H, Zhao Y, Li Y, Tsui KL (2021) An adaptive weight learning-based multitask deep network for continuous blood pressure estimation using Electrocardiogram signals. Sensors 21(5):1595
Fati SM, Muneer A, Akbar NA, Taib SM (2021) A continuous cuffless blood pressure estimation using tree-based pipeline optimization tool. Symmetry 13(4):686
Feurer M, Klein A, Eggensperger K, Springenberg JT, Blum M, Hutter F (2019) Auto-sklearn: Efficient and robust automated machine learning. In: Automated machine learning. Springer, Berlin, pp 113–134
Fong MWK, Ng E, Jian KEZ, Hong TJ (2019) SVR ensemble-based continuous blood pressure prediction using multi-channel photoplethysmogram. Comput Biol Med 113:103392
for the Advancement of Medical Instrumentation A et al (2003) American national standard. Manual, electronic or automated sphygmomanometers. ANSI/AAMI SP10-2002/A1
Forouzanfar M, Dajani HR, Groza VZ, Bolic M, Rajan S (2011) Feature-based neural network approach for oscillometric blood pressure estimation. IEEE Trans Instrum Meas 60(8):2786–2796
Forouzanfar M, Dajani HR, Groza VZ, Bolic M, Rajan S, Batkin I (2015) Oscillometric blood pressure estimation: past, present, and future. IEEE Rev Biomed Eng 8:44–63
Franco G, Cerina L, Gallicchio C, Micheli A, Santambrogio MD (2019) Continuous blood pressure estimation through optimized echo state networks. In: International conference on artificial neural networks (ICANN). Springer, pp 48–61
Fujita D, Suzuki A, Ryu K (2019) PPG-based systolic blood pressure estimation method using PLS and level-crossing feature. Appl Sci 9(2):304
Gama J, Sebastiao R, Rodrigues PP (2009) Issues in evaluation of stream learning algorithms. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining (KDD), pp 329–338
Ganti VG, Carek A, Nevius BN, Heller J, Etemadi M, Inan O (2020) Wearable cuff-less blood pressure estimation at home via pulse transit time. IEEE J Biomed Health Inform
Ganti V, Carek AM, Jung H, Srivatsa AV, Cherry D, Johnson LN, Inan OT et al (2021) Enabling wearable pulse transit time-based blood pressure estimation for medically underserved areas and health equity: Comprehensive evaluation study. JMIR Mhealth Uhealth 9(8):27466
Gao SC, Wittek P, Zhao L, Jiang WJ (2016) Data-driven estimation of blood pressure using photoplethysmographic signals. In: 2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 766–769. IEEE
Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) Physiobank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):215–220
Golino HF, Amaral LSDB, Duarte SFP, Gomes CMA, Soares TDJ, Reis LAS, Santos J (2014) Predicting increased blood pressure using machine learning. J Obes
Gonzalez Viejo C, Fuentes S, Torrico DD, Dunshea FR (2018) Non-contact heart rate and blood pressure estimations from video analysis and machine learning modelling applied to food sensory responses: A case study for chocolate. Sensors 18(6):1802
Goyal P, Dollár P, Girshick R, Noordhuis P, Wesolowski L, Kyrola A, Tulloch A, Jia Y, He K (2017) Accurate, large minibatch SGD: training ImageNet in 1 hour. arXiv preprint arXiv:1706.02677
Ha S, Choi S (2016)Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In: 2016 international joint conference on neural networks (IJCNN), pp 381–388. IEEE
Haddad S, Boukhayma A, Caizzone A (2021) Continuous PPG-based blood pressure monitoring using multi-linear regression. IEEE J Biomed Health Inform 26(5):2096–2105
Hakak S, Ray S, Khan WZ, Scheme E (2020) A framework for edge-assisted healthcare data analytics using federated learning. In: 2020 IEEE international conference on big data (IEEE Big Data), pp 3423–3427. IEEE

Harfiya LN, Chang C-C, Li Y-H (2021) Continuous blood pressure estimation using exclusively photoplethysmography by LSTM-based signal-to-signal translation. Sensors 21(9):2952
Hasanzadeh N, Ahmadi MM, Mohammadzade H (2019) Blood pressure estimation using photoplethysmogram signal and its morphological features. IEEE Sens J 20(8):4300–4310
Hassani A, Foruzan AH (2019) Improved PPG-based estimation of the blood pressure using latent space features. Signal Image Video Process 13(6):1141–1147
He H, Chen S, Li K, Xu X (2011) Incremental learning from stream data. IEEE Trans Neural Netw 22(12):1901–1914
He R, Huang Z-P, Ji, L-Y, Wu J-K, Li H, Zhang Z-Q (2016a) Beat-to-beat ambulatory blood pressure estimation based on random forest. In: 2016 IEEE 13th international conference on wearable and implantable body sensor networks (BSN), pp 194–198. IEEE
He K, Zhang X, Ren S, Sun J (2016b) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
He X, Zhao K, Chu X (2021) AutoML: a survey of the state-of-the-art. Knowl Based Syst 212:106622
Hill BL, Rakocz N, Rudas Á, Chiang JN, Wang S, Hofer I, Cannesson M, Halperin E (2021) Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning. Sci Rep 11(1):1–12
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
Hoi SCH, Sahoo D, Lu J, Zhao P (2021) Online learning: a comprehensive survey. Neurocomputing 459:249–289
Hosanee M, Chan G, Welykholowa K, Cooper R, Kyriacou PA, Zheng D, Allen J, Abbott D, Menon C, Lovell NH et al (2020) Cuffless single-site photoplethysmography for blood pressure monitoring. J Clin Med 9(3):723
Hsieh Y-Y, Wu C-D, Lu S-S, Tsao Y (2016) A linear regression model with dynamic pulse transit time features for noninvasive blood pressure prediction. In: 2016 IEEE biomedical circuits and systems conference (BioCAS), pp 604–607. IEEE
Hsu Y-C, Li Y-H, Chang C-C, Harfiya LN (2020) Generalized deep neural network model for cuffless blood pressure estimation with photoplethysmogram signal only. Sensors 20(19):5668
Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond Ser A 454(1971):903–995
Huang K-H, Tan F, Wang T-D, Yang Y-J (2019) A highly sensitive pressure-sensing array for blood pressure estimation assisted by machine-learning techniques. Sensors 19(4):848
Huang B, Chen W, Lin C-L, Juang C-F, Wang J (2022) MLP-BP: a novel framework for cuffless blood pressure measurement with PPG and ECG signals based on MLP-Mixer neural networks. Biomed Signal Process Control 73:103404
Hui X, Conroy TB, Kan EC (2020) Multi-point near-field RF sensing of blood pressures and heartbeat dynamics. IEEE Access 8:89935–89945
Hussain Z, Gimenez F, Yi D, Rubin D (2017) Differential data augmentation techniques for medical imaging classification tasks. In: AMIA annual symposium proceedings, vol 2017. American Medical Informatics Association, p 979
Huttunen JM, Kärkkäinen L, Lindholm H (2019) Pulse transit time estimation of aortic pulse wave velocity and blood pressure using machine learning and simulated training data. PLoS Comput Biol 15(8):1007259
Ibrahim B, Jafari R (2019) Cuffless blood pressure monitoring from an array of wrist bio-impedance sensors using subject-specific regression models: Proof of concept. IEEE Trans Biomed Circ Syst 13(6):1723–1735
Ibtehaz N, Rahman MS (2020) PPG2ABP: Translating photoplethysmogram (PPG) signals to arterial blood pressure (ABP) waveforms using fully convolutional neural networks. arXiv preprint arXiv:2005.01669
Jeong IC, Finkelstein J (2016) Introducing contactless blood pressure assessment using a high speed video camera. J Med Syst 40(4):77
Jeong DU, Lim KM (2021) Combined deep CNN-LSTM network-based multitasking learning architecture for noninvasive continuous blood pressure estimation using difference in ECG-PPG features. Sci Rep 11(1):1–8
Jeong Y-S, Kang AR, Jung W, Lee SJ, Lee S, Lee M, Chung YH, Koo BS, Kim SH (2019) Prediction of blood pressure after induction of anesthesia using deep learning: a feasibility study. Appl Sci 9(23):5135
Ji J, Dong M, Lin Q, Tan KC (2022) Noninvasive cuffless blood pressure estimation with dendritic neural regression. IEEE Trans Cybern

Johnson AE, Pollard TJ, Shen L, Li-Wei HL, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG (2016) MIMIC-III, a freely accessible critical care database. Sci Data 3(1):1–9
Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R (2021) MIMIC-IV (version 1.0). https://doi.org/10.13026/s6n6-xd98
Joshi SL, Vatti RA, Tornekar RV (2013) A survey on ECG signal denoising techniques. In: 2013 international conference on communication systems and network technologies (CSNT), pp 60–64. IEEE
Kachuee M, Kiani MM, Mohammadzade H, Shabany M (2015) Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. In: 2015 IEEE international symposium on circuits and systems (ISCAS), pp 1006–1009. IEEE
Kachuee M, Kiani MM, Mohammadzade H, Shabany M (2016) Cuffless blood pressure estimation algorithms for continuous health-care monitoring. IEEE Trans Biomed Eng 64(4):859–869
Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95-international conference on Neural Networks (ICNN), vol 4, pp 1942–1948. IEEE
Khaleghi B, Khamis A, Karray FO, Razavi SN (2013) Multisensor data fusion: a review of the state-of-the-art. Inf. Fusion 14(1):28–44
Khalid SG, Zhang J, Chen F, Zheng D (2018) Blood pressure estimation using photoplethysmography only: comparison between different machine learning approaches. J Healthc Eng
Khalid SG, Liu H, Zia T, Zhang J, Chen F, Zheng D (2020) Cuffless blood pressure estimation using single channel photoplethysmography: A two-step method. IEEE Access 8:58146–58154
Khan Mamun MMR, Alouani AT (2022) Cuffless blood pressure measurement using linear and nonlinear optimized feature selection. Diagnostics 12(2):408
Kido K, Chen Z, Huang M, Tamura T, Chen W, Ono N, Takeuchi M, Altaf-Ul-Amin M, Kanaya S (2022) Discussion of cuffless blood pressure prediction using plethysmograph based on a longitudinal experiment: is the individual model necessary? Life 12(1):11
Kissas G, Yang Y, Hwuang E, Witschey W, Detre J, Perdikaris P (2020) Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4d flow MRI data using physics-informed neural networks. Comput Methods Appl Mech Eng 358:112623. https://doi.org/10.1016/j.cma.2019.112623
Koshimizu H, Kojima R, Kario K, Okuno Y (2020) Prediction of blood pressure variability using deep neural networks. Int J Med Inform 136:104067
Krawczyk B (2016) Learning from imbalanced data: Open challenges and future directions. Prog Artif Intell 5(4):221–232
Kurylyak Y, Barbé K, Lamonaca F, Grimaldi D, Van Moer W (2013) Photoplethysmogram-based blood pressure evaluation using kalman filtering and neural networks. In: 2013 IEEE international symposium on medical measurements and applications (MeMeA), pp 170–174. IEEE
Landry C, Peterson SD, Arami A (2019) Estimation of the blood pressure waveform using Electrocardiography. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 7060–7063. IEEE
Lazazzera R, Belhaj Y, Carrault G (2019) A new wearable device for blood pressure estimation using photoplethysmogram. Sensors 19(11):2557
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
LeDell E, Poirier S (2020) H2o AutoML: scalable automatic machine learning. In: Proceedings of the AutoML workshop at ICML (ICML Workshop)
Lee S, Chang J-H (2016) Oscillometric blood pressure estimation based on deep learning. IEEE Trans Ind Inf 13(2):461–472
Lee S, Chang J-H (2017a) Deep belief networks ensemble for blood pressure estimation. IEEE Access 5:9962–9972
Lee S, Chang J-H (2017b) Deep Boltzmann regression with mimic features for oscillometric blood pressure estimation. IEEE Sens J 17(18):5982–5993
Lee S, Chang J-H (2019) Dempster-Shafer fusion based on a deep boltzmann machine for blood pressure estimation. Appl Sci 9(1):96
Lee H-C, Jung C-W (2018) Vital recorder-a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices. Sci Rep 8(1):1–8
Lee S, Lee G (2020) Ensemble methodology for confidence interval in oscillometric blood pressure measurements. J Med Syst 44(5):1–9
Lee S, Ahmad A, Jeon G (2018) Combining bootstrap aggregation with support vector regression for small blood pressure measurement. J Med Syst 42(4):1–7
Lee S, Lee G, Jeon G (2019a) Statistical approaches based on deep learning regression for verification of normality of blood pressure estimates. Sensors 19(9):2137

Lee J, Yang S, Lee S, Kim HC (2019b) Analysis of pulse arrival time as an indicator of blood pressure in a large surgical biosignal database: recommendations for developing ubiquitous blood pressure monitoring methods. J Clin Med 8(11):1773
Lee S, Dajani HR, Rajan S, Lee G, Groza VZ (2020) Uncertainty in blood pressure measurement estimated using ensemble-based recursive methodology. Sensors 20(7):2108
Lee D, Kwon H, Son D, Eom H, Park C, Lim Y, Seo C, Park K (2021) Beat-to-beat continuous blood pressure estimation using bidirectional long short-term memory network. Sensors 21(1):96
Leitner JJ, Chiang P-H, Dey S (2021) Personalized blood pressure estimation using photoplethysmography: a transfer learning approach. IEEE J Biomed Health Inform 26(1):218–228
Li Z, He W (2021) A continuous blood pressure estimation method using photoplethysmography by GRNN-based model. Sensors 21(21):7207
Li P, Laleg-Kirati T-M (2021) Central blood pressure estimation from distal PPG measurement using semi-classical signal analysis features. IEEE Access 9:44963–44973
Li C, Zheng C, Tai C (1995) Detection of ECG characteristic points using wavelet transforms. IEEE Trans Biomed Eng 42(1):21–28
Li X, Wu S, Wang L (2017) Blood pressure prediction via recurrent models with contextual layer. In: Proceedings of the 26th international conference on world wide web (WWW), pp 685–693
Li Y-H, Harfiya LN, Purwandari K, Lin Y-D (2020a) Real-time cuffless continuous blood pressure estimation using deep learning model. Sensors 20(19):5606
Li T, Sahu AK, Talwalkar A, Smith V (2020b) Federated learning: challenges, methods, and future directions. IEEE Signal Process Mag 37(3):50–60
Li Y-H, Harfiya LN, Chang C-C (2021) Featureless blood pressure estimation based on photoplethysmography signal using CNN and BiLSTM for IoT devices. Wirel Commun Mob Comput
Liang Y, Elgendi M, Chen Z, Ward R (2018) An optimal filter for short photoplethysmogram signals. Sci Data 5(1):1–12
Liang Y, Chen Z, Liu G, Elgendi M (2018) A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in china. Sci Data 5(1):1–7
Liang W, Tadesse GA, Ho D, Li F-F, Zaharia M, Zhang C, Zou J (2022) Advances, challenges and opportunities in creating data for trustworthy AI. Nat Mach Intell 4:669–677. https://doi.org/10.1038/s42256-022-00516-1
Lin W-H, Li X, Li Y, Li G, Chen F (2020) Investigating the physiological mechanisms of the photoplethysmogram features for blood pressure estimation. Physiol Meas 41(4):044003
Lin W-H, Chen F, Geng Y, Ji N, Fang P, Li G (2021a) Towards accurate estimation of cuffless and continuous blood pressure using multi-order derivative and multivariate photoplethysmogram features. Biomed Signal Process Control 63:102198
Lin W, Demirel BU, Al Faruque MA, Li G (2021b) Energy-efficient blood pressure monitoring based on single-site photoplethysmogram on wearable devices. In: 2021 43rd annual international conference of the IEEE engineering in medicine & biology society (EMBC), pp 504–507. IEEE
Liu D, Görges M, Jenkins SA (2012) University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research. Anesth Analg 114(3):584–589
Liu M, Po L-M, Fu H (2017) Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative. Int J Comput Theory Eng 9(3):202
Liu J, Wu Y, Yuan Z, Sun X (2018) Blood pressure prediction with multi-cue based RBF and LSTM model. In: 2018 9th international conference on information technology in medicine and education (ITME), pp 72–76. IEEE
Liu J, Qiu S, Luo N, Lau S-K, Yu H, Kwok T, Zhang Y-T, Zhao N (2020a) PCA-based multi-wavelength photoplethysmography algorithm for cuffless blood pressure measurement on elderly subjects. IEEE J Biomed Health Inform 25(3):663–673
Liu Z, Zhou B, Li Y, Tang M, Miao F (2020b) Continuous blood pressure estimation from electrocardiogram and photoplethysmogram during arrhythmias. Front Physiol 11:1126
Liu Q, Zheng Y, Zhang Y, Poon CC (2021) Beats-to-beats estimation of blood pressure during supine cycling exercise using a probabilistic nonparametric method. IEEE Access 9:115655–115663
Liwicki M, Graves A, Fernàndez S, Bunke H, Schmidhuber J (2007) A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks. In: Proceedings of the 9th international conference on document analysis and recognition (ICDAR)
Lo FP-W, Li CX-T, Wang J, Cheng J, Meng MQ-H (2017) Continuous systolic and diastolic blood pressure estimation utilizing long short-term memory network. In: 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 1853–1856. IEEE

Luo H, Yang D, Barszczyk A, Vempala N, Wei J, Wu SJ, Zheng PP, Fu G, Lee K, Feng Z-P (2019) Smartphone-based blood pressure measurement using transdermal optical imaging technology. Circulation 12(8):008857

Ma Y, Choi J, Hourlier-Fargette A, Xue Y, Chung HU, Lee JY, Wang X, Xie Z, Kang D, Wang H, Han S, Kang S-K, Kang Y, Yu X, Slepian M, Raj M, Model J, Feng X, Ghaffari R, Huang Y-S (2018) Relation between blood pressure and pulse wave velocity for human arteries. Proc Natl Acad Sci USA 115:201814392. https://doi.org/10.1073/pnas.1814392115

MacMahon S, Peto R, Collins R, Godwin J, Cutler J, Sorlie P, Abbott R, Neaton J, Dyer A, Stamler J (1990) Blood pressure, stroke, and coronary heart disease: part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. The Lancet 335(8692):765–774

Magbool A, Bahloul MA, Ballal T, Al-Naffouri TY, Laleg-Kirati T-M (2021) Aortic blood pressure estimation: a hybrid machine-learning and cross-relation approach. Biomed Signal Process Control 68:102762

Mahajan S (2021) Vayu: an open-source toolbox for visualization and analysis of crowd-sourced sensor data. Sensors 21(22):7726

Maher N, Elsheikh G, Anis W, Emara T (2021) Enhancement of blood pressure estimation method via machine learning. Alex Eng J 60(6):5779–5796

Mahmud S, Ibtehaz N, Khandakar A, Tahir A, Rahman T, Islam K, Hossain S, Rahman M, Musharavati F, Ayari M, Islam M, Chowdhury M (2022) A shallow U-Net architecture for reliably predicting blood pressure (BP) from photoplethysmogram (PPG) and electrocardiogram (ECG) signals. Sensors 22:919. https://doi.org/10.3390/s22030919

Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA (2021) NeuroKit2: a python toolbox for neurophysiological signal processing. Behav Res Methods 53(4):1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Malayeri AB, Khodabakhshi MB (2022) Concatenated convolutional neural network model for cuffless blood pressure estimation using fuzzy recurrence properties of PPG signals

Maqsood S, Xu S, Tran S, Garg S, Springer M, Karunanithi M, Mohawesh R (2022) A survey: from shallow to deep machine learning approaches for blood pressure estimation using biosensors. Expert Syst Appl 116788

Martinez-Ríos E, Montesinos L, Alfaro-Ponce M, Pecchia L (2021) A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data. Biomed Signal Process Control 68:102813

Marzorati D, Bovio D, Salito C, Mainardi L, Cerveri P (2020) Chest wearable apparatus for cuffless continuous blood pressure measurements based on PPG and PCG signals. IEEE Access 8:55424–55437

Matsumura K, Rolfe P, Toda S, Yamakoshi T (2018) Cuffless blood pressure estimation using only a smartphone. Sci Rep 8(1):1–9

Mehrabadi MA, Aqajari SAH, Zargari AHA, Dutt N, Rahmani AM (2022) Novel blood pressure waveform reconstruction from photoplethysmography using cycle generative adversarial networks. arXiv preprint arXiv:2201.09976

Miao F, Fu N, Zhang Y-T, Ding X-R, Hong X, He Q, Li Y (2017) A novel continuous blood pressure estimation approach based on data mining techniques. IEEE J Biomed Health Inform 21(6):1730–1740

Miao F, Liu Z-D, Liu J-K, Wen B, He Q-Y, Li Y (2019) Multi-sensor fusion approach for cuff-less blood pressure measurement. IEEE J Biomed Health Inform 24(1):79–91

Miao F, Wen B, Hu Z, Fortino G, Wang X-P, Liu Z-D, Tang M, Li Y (2020) Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques. Artif Intell Med 108:101919

Microsoft: a Fast Library for Automated Machine Learning & Tuning (2021). https://github.com/microsoft/FLAML

Mishra B, Nirala NS (2020) A survey on denoising techniques of PPG signal. In: 2020 IEEE international conference for innovation in technology (INOCON), pp 1–8. IEEE

MIT-LCP: WFDB-python. https://github.com/MIT-LCP/wfdb-python

Monte-Moreno E (2011) Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. Artif Intell Med 53(2):127–138

Moody GB, Mark RG (1996) A database to support development and evaluation of intelligent intensive care monitoring. In: Computers in cardiology (CIC). IEEE, pp 657–660

Moss L, Corsar D, Shaw M, Piper I, Hawthorne C (2022) Demystifying the black box: the importance of interpretability of predictive models in neurocritical care. Neurocrit Care. https://doi.org/10.1007/s12028-022-01504-4

Mousavi SS, Charmi M, Firouzmand M, Hemmati M, Moghadam M (2019a) A new approach based on dynamical model of the ECG signal to blood pressure estimation. In: 2019 4th international conference on pattern recognition and image analysis (IPRIA), pp 210–215. IEEE

Mousavi SS, Firouzmand M, Charmi M, Hemmati M, Moghadam M, Ghorbani Y (2019b) Blood pressure estimation from appropriate and inappropriate PPG signals using a whole-based method. Biomed Signal Process Control 47:196–206

Mousavi SS, Charmi M, Firouzmand M, Hemmati M, Moghadam M, Ghorbani Y (2020) ECG-based blood pressure estimation using Mechano-Electric coupling concept. arXiv preprint arXiv:2008.10099

Mukkamala R, Hahn J-O, Inan OT, Mestha LK, Kim C-S, Töreyin H, Kyal S (2015) Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice. IEEE Trans Biomed Eng 62(8):1879–1901

Mukkamala R, Yavarimanesh M, Natarajan K, Hahn J-O, Kyriakoulis KG, Avolio AP, Stergiou GS (2021) Evaluation of the accuracy of cuffless blood pressure measurement devices: Challenges and proposals. Hypertension 78(5):1161–1167

Musini VM, Wright JM (2009) Factors affecting blood pressure variability: Lessons learned from two systematic reviews of randomized controlled trials. PLoS ONE 4(5):5673

Natarajan K, Block RC, Yavarimanesh M, Chandrasekhar A, Mestha LK, Inan O, Hahn J-O, Mukkamala R (2021) Photoplethysmography fast upstroke time intervals can be useful features for cuff-less measurement of blood pressure changes in humans. IEEE Trans Biomed Eng

Nemenyi PB (1963) Distribution-free multiple comparisons. Princeton University, New Jersey

Norgaard M, Ravn O, Poulsen NK, Hansen LK (2000) Neural networks for modelling and control of dynamic systems: a practitioner's handbook. Springer, Berlin

O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, Altman D, Bland M, Coats A, Atkins N et al (1993) The British Hypertension Society protocol for the evaluation of blood pressure measuring devices. J Hypertens 11(Suppl 2):43–62

Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359

Pan F, He P, Chen F, Zhang J, Wang H, Zheng D (2019) A novel deep learning based automatic auscultatory method to measure blood pressure. Int J Med Inform 128:71–78

Panwar M, Gautam A, Biswas D, Acharyya A (2020) PP-Net: a deep learning framework for PPG-based blood pressure and heart rate estimation. IEEE Sens J 20(17):10000–10011

Park DY, Joe D, Kim DH, Park H, Han JH, Jeong CK, Park H, Park J, Joung B, Lee K (2017) Self-powered real-time arterial pulse monitoring using ultrathin epidermal piezoelectric sensors. Adv Mater. https://doi.org/10.1002/adma.201702308

Paviglianiti A, Randazzo V, Cirrincione G, Pasero E (2020a) Double channel neural non invasive blood pressure prediction. In: International conference on intelligent computing (ICIC). Springer, New York, pp 160–171

Paviglianiti A, Randazzo V, Cirrincione G, Pasero E (2020b) Neural recurrent approches to noninvasive blood pressure estimation. In: 2020 international joint conference on neural networks (IJCNN), pp 1–7. IEEE

Peng R-C, Yan W-R, Zhang N-L, Lin W-H, Zhou X-L, Zhang Y-T (2015) Cuffless and continuous blood pressure estimation from the heart sound signals. Sensors 15(9):23653–23666

Peter L, Noury N, Cerny M (2014) A review of methods for non-invasive and continuous blood pressure monitoring: pulse transit time method is promising? IRBM 35(5):271–282

Phillips AZ, Kiefe CI, Lewis CE, Schreiner PJ, Tajeu GS, Carnethon MR (2022) Alcohol use and blood pressure among adults with hypertension: the mediating roles of health behaviors. J Gen Intern Med 1–8

Pickering TG, Harshfield GA, Kleinert HD, Blank S, Laragh JH (1982) Blood pressure during normal daily activities, sleep, and exercise: comparison of values in normal and hypertensive subjects. J Am Med Assoc 247(7):992–996

Picone DS, Schultz MG, Otahal P, Aakhus S, Al-Jumaily AM, Black JA, Bos WJ, Chambers JB, Chen C-H, Cheng H-M et al (2017) Accuracy of cuff-measured blood pressure: systematic reviews and meta-analyses. J Am Coll Cardiol 70(5):572–586

Qin K, Huang W, Zhang T (2021) Deep generative model with domain adversarial training for predicting arterial blood pressure waveform from photoplethysmogram signal. Biomed Signal Process Control 70:102972

Qiu Y, Liu D, Yang G, Qi D, Lu Y, He Q, Qian X, Li X, Cao Y, Shuai J (2021) Cuffless blood pressure estimation based on composite neural network and graphics information. Biomed Signal Process Control 70:103001

Radha M, De Groot K, Rajani N, Wong CC, Kobold N, Vos V, Fonseca P, Mastellos N, Wark PA, Velthoven N et al (2019) Estimating blood pressure trends and the nocturnal dip from photoplethysmography. Physiol Meas 40(2):025006

Ramakrishnan A, Prathosh A, Ananthapadmanabha T (2014) Threshold-independent QRS detection using the dynamic plosion index. IEEE Signal Proc Lett 21(5):554–558

Riaz F, Azad MA, Arshad J, Imran M, Hassan A, Rehman S (2019) Pervasive blood pressure monitoring using Photoplethysmogram (PPG) sensor. Future Gener Comput Syst 98:120–130

Rong M, Li K (2021a) A multi-type features fusion neural network for blood pressure prediction based on photoplethysmography. Biomed Signal Process Control 68:102772

Rong M, Li K (2021b) A blood pressure prediction method based on imaging photoplethysmography in combination with machine learning. Biomed Signal Process Control 64:102328

Ruder S (2017) An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098

Sadrawi M, Shieh J-S, Fan SZ, Lin CH, Haraikawa K, Chien JC, Abbod MF (2016) Intermittent blood pressure prediction via multiscale entropy and ensemble artificial neural networks. In: 2016 IEEE EMBS conference on biomedical engineering and sciences (IECBES), pp 356–359. IEEE

Sadrawi M, Lin Y-T, Lin C-H, Mathunjwa B, Fan S-Z, Abbod MF, Shieh J-S (2020) Genetic deep convolutional autoencoder applied for generative continuous arterial blood pressure via photoplethysmography. Sensors 20(14):3829

Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L-W, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG (2011) Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database. Crit Care Med 39(5):952

Sagirova Z, Kuznetsova N, Gogiberidze N, Gognieva D, Suvorov A, Chomakhidze P, Omboni S, Saner H, Kopylov P (2021) Cuffless blood pressure measurement using a smartphone-case based ECG monitor with photoplethysmography in hypertensive patients. Sensors 21(10):3525

Samartkit P, Pullteap S, Bernal O (2022) A non-invasive heart rate and blood pressure monitoring system using piezoelectric and photoplethysmographic sensors. Measurement 196:111211

Sasso AM, Datta S, Jeitler M, Steckhan N, Kessler CS, Michalsen A, Arnrich B, Böttinger E (2020) HYPE: predicting blood pressure from photoplethysmograms in a hypertensive population. In: International conference on artificial intelligence in medicine (AIME). Springer, pp 325–335

Schlesinger O, Vigderhouse N, Eytan D, Moshe Y (2020) Blood pressure estimation from PPG signals using convolutional neural networks and Siamese network. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 1135–1139 . IEEE

Schlesinger O, Vigderhouse N, Moshe Y, Eytan D (2020) Estimation and tracking of blood pressure using routinely acquired photoplethysmographic signals and deep neural networks. Crit Care Explor 2(4)

Scholkmann F, Boss J, Wolf M (2012) An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals. Algorithms 5(4):588–603

Schrumpf F, Frenzel P, Mönch C, Osterhoff G, Fuchs M (2020) PPG-based blood pressure estimation using residual neural networks and spectrograms. In: 54th annual conference of the German society of biomedical engineering (BMT), p 1

Schrumpf F, Frenzel P, Aust C, Osterhoff G, Fuchs M (2021a) Assessment of non-invasive blood pressure prediction from PPG and rPPG signals using deep learning. Sensors 21(18):6022

Schrumpf F, Frenzel P, Aust C, Osterhoff G, Fuchs M (2021b) Assessment of deep learning based blood pressure prediction from PPG and rPPG signals. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR Workshop), pp 3820–3830

Schwenk H, Bengio Y (2000) Boosting neural networks. Neural Comput 12(8):1869–1887

Secerbegovic A, Bergsland J, Halvorsen PS, Suljanovic N, Mujcic A, Balasingham I (2016) Blood pressure estimation using video plethysmography. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI), pp 461–464 . IEEE

Senturk U, Polat K, Yucedag I (2020) A non-invasive continuous cuffless blood pressure estimation using dynamic recurrent neural networks. Appl Acoust 170:107534

Sethi T, Kalia A, Sharma A, Nagori A (2020) Interpretable artificial intelligence: closing the adoption gap in healthcare, pp 3–29. https://doi.org/10.1016/B978-0-12-817133-2.00001-X

Shao J, Shi P, Hu S, Yu H (2020) A revised point-to-point calibration approach with adaptive errors correction to weaken initial sensitivity of cuff-less blood pressure estimation. Sensors 20(8):2205

Sharifi I, Goudarzi S, Khodabakhshi MB (2019) A novel dynamical approach in continuous cuffless blood pressure estimation based on ECG and PPG signals. Artif Intell Med 97:143–151

Shimazaki S, Bhuiyan S, Kawanaka H, Oguri K (2018) Features extraction for cuffless blood pressure estimation by autoencoder from photoplethysmography. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 2857–2860. IEEE

Shin H, Min SD (2017) Feasibility study for the non-invasive blood pressure estimation based on PPG morphology: normotensive subject study. Biomed Eng Online 16(1):1–14
Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J. Big Data 6(1):1–48
Simjanoska M, Gjoreski M, Gams M, Madevska Bogdanova A (2018) Non-invasive blood pressure estimation from ECG using machine learning techniques. Sensors 18(4):1160
Simjanoska M, Kochev S, Tanevski J, Bogdanova AM, Papa G, Eftimov T (2020) Multi-level information fusion for learning a blood pressure predictive model using sensor data. Inf Fusion 58:24–39
Singh G, Kaur R (2013) Removal of EMG interference from Electrocardiogram using back propagation. Int J Innov Res Comput Commun Eng 1(6):1300–1305
Singla M, Sistla P, Azeemuddin S (2019) Cuff-less blood pressure measurement using supplementary ECG and PPG features extracted through wavelet transformation. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 4628–4631. IEEE
Singla M, Azeemuddin S, Sistla P (2020a) Accurate fiducial point detection using Haar wavelet for beat-by-beat blood pressure estimation. IEEE J Transl Eng Health Med 8:1–11
Singla M, Azeemuddin S, Sistla P (2020b) Learning-based model for central blood pressure estimation using feature extracted from ECG and PPG signals. In: 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC), pp 855–858. IEEE
Slapničar G, Luštrek M, Marinko M (2018) Continuous blood pressure estimation from PPG signal. Informatica 42(1)
Slapničar G, Mlakar N, Luštrek M (2019) Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. Sensors 19(15):3420
Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. Adv Neural Inf Process Systems 25
Sola J, Proença M, Ferrario D, Porchet J-A, Falhi A, Grossenbacher O, Allemann Y, Rimoldi SF, Sartori C (2013) Noninvasive and nonocclusive blood pressure estimation via a chest sensor. IEEE Trans Biomed Eng 60(12):3505–3513
Song K, Chung K-Y, Chang J-H (2019) Cuffless deep learning-based blood pressure estimation for smart wristwatches. IEEE Trans Instrum Meas 69(7):4292–4302
Song K, Park T-J, Chang J-H (2021) Novel data augmentation employing multivariate Gaussian distribution for neural network-based blood pressure estimation. Appl Sci 11(9):3923
sppmg: Peak Detection (2017) https://github.com/sppmg/peakDetection
Steinman J, Barszczyk A, Sun H-S, Lee K, Feng Z-P (2021) Smartphones and video cameras: future methods for blood pressure measurement. Front Digit Health 3
Su P, Ding X-R, Zhang Y-T, Liu J, Miao F, Zhao N (2018) Long-term blood pressure prediction with deep recurrent neural networks. In: 2018 IEEE EMBS international conference on biomedical & health informatics (BHI), pp 323–328. IEEE
Sugita N, Obara K, Yoshizawa M, Abe M, Tanaka A, Homma N (2015) Techniques for estimating blood pressure variation using video images. In: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 4218–4221. IEEE
Takahashi R, Ogawa-Ochiai K, Tsumura N (2020) Non-contact method of blood pressure estimation using only facial video. Artif Life Robot 25(3):343–350
Tamura T (2021) Cuffless blood pressure monitors: principles, standards and approval for medical use. IEICE Trans Commun 104(6):580–586
Tan X, Ji Z, Zhang Y (2018) Non-invasive continuous blood pressure measurement based on mean impact value method, BP neural network, and genetic algorithm. Technol Health Care 26(S1):87–101
Tanveer MS, Hasan MK (2019) Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ANN-LSTM network. Biomed Signal Process Control 51:382–392
Tasli HE, Gudi A, Den Uyl M (2014) Remote PPG based vital sign measurement using adaptive facial regions. In: 2014 IEEE international conference on image processing (ICIP), pp 1410–1414. IEEE
Tazarv A, Levorato M (2021) A deep learning approach to predict blood pressure from PPG signals. In: 2021 43rd Annual international conference of the IEEE engineering in medicine & biology society (EMBC), pp 5658–5662. IEEE
Thambiraj G, Gandhi U, Devanand V, Mangalanathan U (2019) Noninvasive cuffless blood pressure estimation using pulse transit time, Womersley number, and photoplethysmogram intensity ratio. Physiol Meas 40:075001. https://doi.org/10.1088/1361-6579/ab1f17
Thambiraj G, Gandhi U, Mangalanathan U, Jose VJM, Anand M (2020) Investigation on the effect of Womersley number, ECG and PPG features for cuff less blood pressure estimation using machine learning. Biomed Signal Process Control 60:101942

Tjahjadi H, Ramli K, Murfi H (2020) Noninvasive classification of blood pressure based on photoplethysmography signals using bidirectional long short-term memory and time-frequency analysis. IEEE Access 8:20735–20748
Vanschoren J (2018) Meta-learning: a survey. arXiv preprint arXiv:1810.03548
Vijayarangan S, Vignesh R, Murugesan B, Preejith S, Joseph J, Sivaprakasam M (2020) RPnet: a deep learning approach for robust R peak detection in noisy ECG. In: 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC), pp 345–348. IEEE
Vilalta R, Drissi Y (2002) A perspective view and survey of meta-learning. Artif Intell Rev 18(2):77–95
Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17(3):261–272
Visvanathan A, Banerjee R, Dutta Choudhury A, Sinha A, Kundu S (2014) Smart phone based blood pressure indicator. In: Proceedings of the 4th ACM MobiHoc workshop on pervasive wireless healthcare (MobileHealth), pp 19–24
Wang T-W, Lin S-F (2020) Wearable piezoelectric-based system for continuous beat-to-beat blood pressure measurement. Sensors 20(3):851
Wang Z, Zhang Y (2017) A novel frequency domain method for estimating blood pressure from photoplethysmogram. In: Proceedings of the 9th international conference on signal processing systems (ICSPS), pp 201–206
Wang Y, Si Y, Liu L, Zhang J (2017) Continuous blood pressure prediction using pulse features and Elman neural networks. In: 2017 IEEE 17th international conference on communication technology (ICCT), pp 2008–2013. IEEE
Wang Q, Xu Y, Zeng G, Sun M (2018a) Continuous blood pressure estimation based on two-domain fusion model. Comput Math Method Med
Wang L, Zhou W, Xing Y, Zhou X (2018b) A novel neural network model for blood pressure estimation using photoplethesmography without electrocardiogram. J Healthc Eng
Wang W, Zhu L, Marefat F, Mohseni P, Kilgore K, Najafizadeh L (2020) Photoplethysmography-based blood pressure estimation using deep learning. In: 2020 54th Asilomar conference on signals, systems, and computers (ACSSC), pp 945–949 . IEEE
Wang H, Wang Z, Wang P, Yu M, Xu J, Zhang G (2021) A novel approach to estimate blood pressure of blood loss continuously based on stacked auto-encoder neural networks. Biomed Signal Process Control 69:102853
Wang T-W, Syu J-Y, Chu H-W, Sung Y-L, Chou L, Escott E, Escott O, Lin T-T, Lin S-F (2022) Intelligent bio-impedance system for personalized continuous blood pressure measurement. Biosensors 12(3):150
Waring J, Lindvall C, Umeton R (2020) Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artif Intell Med 104:101822
Welch G, Bishop G et al (1995) An introduction to the Kalman filter
Whitley D (1994) A genetic algorithm tutorial. Stat Comput 4(2):65–85
Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adaptive Data Anal 1(01):1–41
Wu C-M, Chuang CY, Chen Y-J, Chen S-C (2016) A new estimate technology of non-invasive continuous blood pressure measurement based on electrocardiograph. Adv Mech Eng 8(6):1687814016653689
Xiang Y, Li S, Zhang P (2021) An exploration in remote blood pressure management: application of daily routine pattern based on mobile data in health management. Fundam Res
Xing X, Sun M (2016) Optical blood pressure estimation with photoplethysmography and FFT-based neural networks. Biomed Opt Express 7(8):3007–3020
Xing X, Ma Z, Zhang M, Zhou Y, Dong W, Song M (2019) An unobtrusive and calibration-free blood pressure estimation method using photoplethysmography and biometrics. Sci Rep 9(1):1–8
Yamakoshi T, Rolfe P, Yamakoshi K-I (2021) Cuffless blood pressure estimation based on haemodynamic principles: progress towards mobile healthcare. PeerJ 9:11479
Yamanaka S, Morikawa K, Morita H, Huh JY, Yamamura O (2021) Calibration-free cuffless blood pressure estimation based on a population with a diverse range of age and blood pressure. Front Med Tech 36
Yan W-R, Peng R-C, Zhang Y-T, Ho D (2019) Cuffless continuous blood pressure estimation from pulse morphology of photoplethysmograms. IEEE Access 7:141970–141977
Yang S, Sohn J, Lee S, Lee J, Kim HC (2020a) Estimation and validation of arterial blood pressure using photoplethysmogram morphology features in conjunction with pulse arrival time in large open databases. IEEE J Biomed Health Inform 25(4):1018–1030

Yang J, Liu F, Wang B, Chen C, Church T, Dukes L, Smith JO (2020b) Blood pressure states transition inference based on multi-state Markov model. IEEE J Biomed Health Inform 25(1):237–246
Yang S, Zhang Y, Cho S-Y, Correia R, Morgan SP (2021) Non-invasive cuff-less blood pressure estimation using a hybrid deep learning model. Opt Quantum Electron 53(2):1–20
Yi Z, Liu Z, Li W, Ruan T, Chen X, Liu J, Yang B, Zhang W (2022a) Piezoelectric dynamics of arterial pulse for wearable continuous blood pressure monitoring. Adv Mater. https://doi.org/10.1002/adma.202110291
Yi Z, Zhang W-M, Yang B (2022b) Piezoelectric approaches for wearable continuous blood pressure monitoring: a review. J Micromech Microeng. https://doi.org/10.1088/1361-6439/ac87ba
Yin S, Li G, Luo Y, Lin L (2021) Cuff-less continuous blood pressure measurement based on multiple types of information fusion. Biomed Signal Process Control 68:102549
Yousefian P, Shin S, Mousavi AS, Tivay A, Kim C-S, Mukkamala R, Jang D-G, Ko BH, Lee J, Kwon U-K et al (2020) Pulse transit time-pulse wave analysis fusion based on wearable wrist ballistocardiogram for cuff-less blood pressure trend tracking. IEEE Access 8:138077–138087
Yu Z, Peng W, Li X, Hong X, Zhao G (2019) Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV), pp 151–160
Zhang Y, Wang Z (2017) A hybrid model for blood pressure prediction from a PPG signal based on MIV and GA-BP neural network. In: 2017 13th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), pp 1989–1993. IEEE
Zhang Q, Zhou D, Zeng X (2017) Highly wearable cuff-less blood pressure and heart rate monitoring with single-arm electrocardiogram and photoplethysmogram signals. Biomed Eng Online 16(1):1–20
Zhang B, Wei Z, Ren J, Cheng Y, Zheng Z (2018) An empirical study on predicting blood pressure using classification and regression trees. IEEE Access 6:21758–21768
Zhang B, Ren J, Cheng Y, Wang B, Wei Z (2019a) Health data driven on continuous blood pressure prediction based on gradient boosting decision tree algorithm. IEEE Access 7:32423–32433
Zhang B, Ren H, Huang G, Cheng Y, Hu C (2019b) Predicting blood pressure from physiological index data using the SVR algorithm. BMC Bioinform 20(1):1–15
Zhang B, Zhu J, Su H (2020a) Toward the third generation of artificial intelligence. Sci Sin Inform 50(9):1281–1302. https://doi.org/10.1360/SSI-2020-0204. (**in chinese**)
Zhang L, Hurley NC, Ibrahim B, Spatz E, Krumholz HM, Jafari R, Bobak MJ (2020b) Developing personalized models of blood pressure estimation from wearable sensors data using minimally-trained domain adversarial neural networks. In: Machine learning for healthcare conference (MLHC), pp 97–120. PMLR
Zhang Y, Zhou C, Huang Z, Ye X (2021a) Study of cuffless blood pressure estimation method based on multiple physiological parameters. Physiol Meas 42(5):055004
Zhang G, Wang Z, Hou F, Wan Z, Chen F, Yu M, Wang J, Wang H (2021b) Heart rate variability enhances the accuracy of non-invasive continuous blood pressure estimation under blood loss. Rev Sci Instrum 92(10):105106
Zhang Y, Zhang X, Cui P, Li S, Tang J (2021c) Key feature selection and model analysis for blood pressure estimation from electrocardiogram, ballistocardiogram and photoplethysmogram. IEEE Access 9:54350–54359
Zheng J, Yu Z (2021) A novel machine learning-based systolic blood pressure predicting model. J Nanomater
Zhou Y, Ni H, Zhang Q, Wu Q (2019) The noninvasive blood pressure measurement based on facial images processing. IEEE Sens J 19(22):10624–10634

Authors and Affiliations

Keke Qin[1] · Wu Huang[2] · Tao Zhang[1] · Shiqi Tang[3]

Keke Qin
keke.qin@tme.com.cn

Tao Zhang
zhangtao@tme.com.cn

Shiqi Tang
tttpgs@163.com

[1] Chengdu Techman Software Co., Ltd, Damian Street, Longquanyi District, Chengdu 610000, Sichuan, China

[2] School of Computer Science, Sichuan University, Chengdu 610000, Sichuan, China

[3] College of Computer and Information Science, Hunan Institute of Technology, Henghua Road, Zhuhui District, Hengyang 421002, Hunan, China