

Wearable Blood Pressure Monitoring Devices: Understanding Heterogeneity in Design and Evaluation

Matt Y. Cheung, Ashutosh Sabharwal, Gerard L. Coté, Ashok Veeraraghavan

Abstract— Objective: Rapid advances in cuffless blood pressure (BP) monitoring over the last decade have the potential to radically transform clinical care for cardiovascular health. However, due to the large heterogeneity in device design and evaluation, it is difficult to critically and quantitatively evaluate research progress in cuffless BP monitoring. In this two-part manuscript, we seek to provide a principled way of describing and accounting for the heterogeneity in device and study design. **Methods:** We first provide an overview of foundational elements and design principles of three critical aspects in the pipeline: 1) sensors and systems, 2) pre-processing and feature extraction, and 3) BP estimation algorithms. Then, we critically analyze the state-of-the-art methods via a systematic review. **Results:** We find a large amount of heterogeneity in study designs making fair comparisons challenging. In addition, many study designs lead to data leakage, and underpowered studies. We suggest a first open-contribution BP estimation benchmark based on existing public datasets for standardized algorithmic comparisons. Second, we observe that BP distribution in the study sample and the time between calibration and test in emerging personalized devices are significant confounders in BP estimation error. We suggest accounting for these using a metric “explained deviation” which is closely related to the coefficient of determination (R^2 , a frequently used statistic). Finally, we complement this manuscript with a website, <https://wearablebp.github.io>, containing a bibliography, meta-analysis results, datasets, and benchmarks, providing a timely platform to understand the state-of-the-art devices. **Conclusion:** There is large heterogeneity in device and study design, which should be carefully accounted for when designing, comparing, and contrasting studies. **Significance:** Our findings will allow readers to parse out the heterogeneous literature and move toward promising directions for safer and more reliable devices in clinical practice and beyond.

Index Terms— Cuffless Blood Pressure, Wearables, Meta-analysis

I. INTRODUCTION

Blood pressure is one of the most important clinical variables for monitoring cardiovascular health. Hypertension (high blood pressure) is estimated to result in 7.5 million deaths annually (~13% of all deaths) [1]. In the United States, CDC estimates that 48% of adults over 18 years of age have hypertension [2]. The development of the first electronic blood pressure monitor in 1973 and the subsequent adoption of their use over the last two decades have significantly impacted clinical practice and clinical guidelines and have improved patient outcomes [3]. Despite these significant impacts of home-based BP monitoring, two unresolved concerns remain for these electronic

Submission Date: 10/28/23. The authors would like to acknowledge support from NSF Expeditions IIS-1730574, NSF PATHS-UP ERC EEC-1648451, and a fellowship from the Gulf Coast Consortia on the NLM Training Program in Biomedical Informatics and Data Science T15LM007093.

M. Y. Cheung is with the Department of Electrical & Computer Engineering at Rice University, Houston, TX 77005 USA. (mattyc@rice.edu).

A. Sabharwal is with the Department of Electrical & Computer Engineering at Rice University, Houston, TX 77005 USA. (ashu@rice.edu).

G. L. Coté is with the Department of Biomedical Engineering at Texas A&M, College Station, TX 77843 USA. (gcote@tamu.edu).

A. Veeraraghavan is with the Department of Electrical & Computer Engineering and Computer Science at Rice University, Houston, TX 77005 USA. (vashok@rice.edu).

cuff-based devices. First, patient compliance with home BP measurement protocols is highly varied, resulting in challenges in clinical decision-making [3], [4]. Second, activity-induced, diurnal, and sleep-time variations in blood pressure are critical and, for the most part, completely ignored in home BP measurement protocols [5], [6].

Cuffless blood pressure measurement devices that can continuously monitor a subject’s blood pressure in the background while they participate in daily life activities provide an opportunity to address the above-mentioned limitations. Over the last decade, tremendous progress has been made in cuffless blood pressure monitoring to the extent that a large fraction of practitioners and experts believe that it is likely that these devices will achieve a level of maturity and reliability that allows the integration of cuffless continuous blood pressure monitoring into clinical practice over the next decade. Despite this promise, the academic and clinical literature around cuffless blood pressure monitoring devices and their evaluation remains highly scattered across multiple disciplines, heterogeneous in study designs, and suffers from meaningful comparisons across different solutions.

There are varying opinions on how to organize and present the vast and fast emerging research progress in this area. One strategy is to present a review report that studies progress in a narrative manner [7] or through the lens of different calibration methods [8]. While some existing reviews focus on specific sub-domains within this broad area, such as sensors (e.g., photoplethysmography [9]–[12]) or features (e.g., pulse transit time [13]–[17]), others delve into singular components of the monitoring process, like algorithms [18], [19]. In spite of this, there are no current reviews that holistically present the various integral elements in cuffless BP monitoring and provide a comparative review that is comprehensive. To this end, we present the first comprehensive review that addresses all elements in cuffless BP monitoring, including designing sensors and systems, pre-processing and feature extraction algorithms, and estimation algorithms. We believe this approach places the reader in a much better position to (i) explore the design space and (ii) understand existing and emerging technologies. We note that most existing reviews (especially the narrative reviews) do not address a central element around evaluating and understanding performance – how does one meaningfully compare metrics across studies when the studies use (a) varying study designs and (b) varying datasets for performance quantification?

Most existing studies do not have external validation, use different calibration techniques, and often do not recruit study populations aligned with existing validation standards. These issues make it difficult for the community to determine and compare across study results, evaluate truly promising research directions, and address shortcomings of current work. Recent commentaries and standards have recognized these issues and have begun articulating “what’s wrong” (i.e., visual representation of flawed results [13], [14]) and “what’s right” (i.e., meeting and redefining the standards) separately; our work can be interpreted in that context. To tackle the lack of external validation, we suggest a first open-contribution BP estimation benchmark based on public datasets for more reliable algorithm comparisons. To tackle heterogeneity in study design and evaluation, we perform a systematic review to motivate a new metric, “explained deviation,” and account for systematic variations in BP distribution and calibration techniques across studies. Explained deviation allows

us to make more reliable comparisons between different devices and populations on a continuum based on reported summary statistics. We want to emphasize that our intention is not to override the established standards but to provide more clarity while evaluating approaches with varying study designs.

Finally, we supplement this review paper with a website, <https://wearablebp.github.io>, containing a bibliography, meta-analysis results, datasets, benchmarks, and available code to provide a timely platform to understand the state of wearable blood pressure monitoring devices.

II. SENSORS AND SYSTEMS

Existing cuff-based BP systems exploit the direct or indirect effects of applying contact pressure on the arteries and blood vessels in the arm to measure BP. However, they are not suitable for wearable, always-on operation. Cuffless devices require passive measurements instead of mechanically interacting with the system by applying pressure and observing the response. The shift from active to passive measurements results in a challenging problem in identifying and measuring surrogate biomarkers that can accurately represent BP. These biomarkers are downstream features that reflect changes in blood pressure, flow, and volume waveforms. Changes in waveforms can be measured using different sensing modalities based on light, impedance, sound, and motion (Fig. 1.). Light-based modalities measure volume changes by illumination to transduce light scattered and absorbed by the tissue. Impedance-based modalities measure volume changes by injecting high-frequency alternating current into the tissue to transduce changes in conductance. Sound-based modalities measure volume changes and velocity by sending ultrasonic waves to transduce the difference in acoustic impedance between tissue boundaries and frequency shifts due to the Doppler effect. Motion-based modalities measure volume and pressure changes passively by transducing applied force from tissue on the sensor. In this section, we review the modalities mentioned above and system-level extensions to increase capabilities for BP estimation.

A. Light-based modalities

Principles. Light-based modalities rely on scattering to detect changes in blood volume in the tissue. The most prevalent modality is photoplethysmography (PPG), where the light source illuminates the tissue, and the photodetector measures the change in light intensity, either reflected or transmitted, from and through the skin, resulting from the absorption and scattering of skin constituents [20]. PPG signals constitute physiological changes (i.e., respiration, venous pulsation, site of measurement, and skin temperature) and individual patients (i.e., skin tone, obesity, age, and gender), and external perturbations (i.e., motion artifacts, ambient light, and contact pressure) [21].

Design choices. Designing PPG-based BP systems requires considerations of transmissive or reflective configurations, wavelength, source intensity, and source-detector distance. In the transmissive configuration, the source and detectors are typically separated on opposing sides of the tissue, and the non-absorbed light is measured. In the reflective configuration, the source and detectors are placed on the same side of the tissue, and the reflected light is measured. While transmissive systems have limited measurement sites (i.e., finger or earlobe), they are typically more robust to motion artifacts and pressure disturbances. While many wavelengths have been used for light-based systems, green, red, and near-infrared (NIR) wavelengths are the most popular. Green wavelengths are motivated by the stronger signal-to-noise ratio in reflected light due to better absorptivity from blood [22]–[24]. Red and NIR wavelengths are

motivated by their lower absorption of lipids, collagen, and melanin in tissue, which facilitates deeper penetration depth [20]. While higher source intensity generally results in a higher signal to the photodetector, it can also result in a higher signal-to-noise ratio (SNR) if more light is transmitted or reflected through the pulsing arteries or arterioles. However, at higher intensities, the system will consume more power, the photodetector may get saturated, and the device may be uncomfortable due to localized heat with long-term use. While larger source-detector distances probe deeper into the skin, the optimal source-detector distance depends on wavelength and skin composition [25], [26]. Higher body-mass indices may lead to lower SNR [27].

Future opportunities. Light-based modalities introduce unique limitations. Ambient light could lead to signal saturation, and the data collected will not contain any information about physiology. Opportunities lie in understanding the tradeoff between applied pressure on the skin and blocking ambient light. High melanin content in darker skin tones results in increased absorption and scattering [28], which may require different calibrations and lead to lower performance in downstream tasks. Opportunities lie in developing sensors and systems that account for and perform equally for different skin tones.

B. Impedance-based modalities

Principles. Impedance-based modalities rely on the penetration of low-amplitude and relatively high-frequency electrical currents to detect changes in electrical impedance in response to arterial distensions. The predominant impedance-based modality is Impedance Plethysmography (IPG). In a four-electrode configuration, the current is injected using the outer electrodes, and the voltage is recorded across the area of interest. Since blood is a better conductor than the surrounding tissue, current will travel via the path of lowest resistance along the artery. The signal amplitude is inversely proportional to changes in blood volume [29] and varies based on tissue composition.

Design choices. Designing impedance-based systems requires considerations of electrical characteristics, electrode arrangement, and electrode size. Higher currents yield higher SNR but increase the risk of tissue damage. Frequencies on the order of 10kHz are used since biological materials are poor conductors and have dielectric properties at higher frequencies [30], but have lower penetration depths due to higher attenuation. Decreasing the electrode areas directly increases the electrode-to-skin impedance and can significantly degrade the performance [31]. Ensuring the artery is along the path of current travel and close to the sensing electrode improves SNR [31].

Future Opportunities. Unlike PPG, IPG can attain deeper penetration depths, can separate the contribution from capillary density and arterial distensions, and is skin tone agnostic. IPG has also demonstrated a small form factor and firm skin-electrode contact [31]–[33]. However, they still suffer from noise due to motion artifacts and contact pressure variations, and further introduce electrical-specific limitations, such as requiring current injection, skin-electrode impedance matching, and bio-compatibility. Opportunities to improve robustness and wearability lie in designing electrodes for impedance-matching and bio-compatibility.

C. Sound-based modalities

Principles. Sound-based modalities send longitudinal acoustic waves that interact with the difference in acoustic impedance between tissues and sense the transmitted or reflected signals scattered by the tissues. The predominant sound-based modalities are imaging and Doppler ultrasound, which measure blood volume and flow.

	Light-based	Impedance-based	Sound-based	Motion-based
Waveform	Volume	Volume	Volume and Flow velocity	Pressure
Source	Illumination	Current injection	Insonation	None (passive)
Transduction	Scattering and absorption of tissue	Changes in conductance	Reflections from acoustic impedance differences at boundary	Applied force on sensor
Characteristics	- Susceptible to motion artifact and, contact pressure variations			
	+ Low cost and ubiquitous – Susceptible to variations in skin tone – Ambient light corrupts signals	+ Skin tone agnostic – Requires skin-electrode impedance matching – Requires bio-compatibility	– High cost (imaging) – Large formfactor – Requires precise Insonation angle – Requires impedance matching for good signals	+ Can quantify contact pressure + Low cost – Performance affected by temperature variations
Common Sensor Placements	Finger, wrist, ear lobe, forehead, toe, upper arm	Wrist, finger, chest	Neck, wrist, chest	Wrist, upper arm, neck

Fig. 1. Principles and characteristics of different modalities used to transduce blood volume, velocity, and pressure waveforms. Light-based modalities measure volume changes by illumination to transduce light scattered and absorbed by the tissue. Impedance-based modalities measure volume changes by injecting high-frequency alternating current into the tissue to transduce changes in conductance. Sound-based modalities measure volume changes and velocity by sending ultrasonic waves to transduce the difference in acoustic impedance between tissue boundaries and frequency shifts due to the Doppler effect. Motion-based modalities measure volume and pressure changes passively by transducing applied force from tissue on the sensor.

Ultrasound devices require impedance matching, usually a thick gel-like mixture of water and propylene glycol placed between the probe and the skin. In imaging ultrasound, an array of transducers transmit and receive pulses. The times-of-flight are proportional to twice the distance between the transducer and tissue, and the spatial pattern reflects the structure of the tissue. Blood volume changes can be tracked by measuring arterial diameter. In Doppler ultrasound, a single transducer transmits a continuous sinusoidal wave. The velocity of blood moving towards and away from the transducer results in higher and lower frequency shifts in the reflected signal (the Doppler effect). Recent efforts such as from the Butterfly iQ (Butterfly Network, Inc., Guilford, CT), Lumify (Philips, Amsterdam, Netherlands), and VScan Air (GE HealthCare, Chicago, IL) have made point-of-care ultrasound a reality.

Design choices. Besides deciding between imaging and Doppler configurations, designing sound-based systems requires considerations of frequency, transducer arrangement, beam profile, and source power. Higher frequency results in better axial resolution but has lower penetration depth [34]. 1-D and 2-D arrays are used depending on whether the application requires lateral resolution. More transducers per unit area increase the lateral resolution but are significantly more expensive to manufacture. A narrower beam profile results in better lateral resolution but sacrifices the field of view [34]. Higher power results in higher SNR but risks tissue damage.

Future opportunities. While ultrasound is also skin tone agnostic and can measure volume and velocity, it requires impedance matching, suffers from high cost (2-D devices), has a large form factor, and requires careful consideration of insonation angle and arterial alignment. Impedance-matching gels placed between the probe and the skin may limit feasibility for every day use due to factors such

as skin irritation, messiness, and drying out. Future opportunities lie in developing impedance-matching gels for long-term every day use. 2-D ultrasound devices can cost thousands of dollars due to the precise manufacturing of transducers en masse. Doppler ultrasound tends to be more affordable. 2-D ultrasound devices also require heavy computation to convert images to physiological signals [35], [36], which may be beyond the limits of current wearable devices. Future opportunities lie in developing devices that leverage efficient algorithms and less resource-intensive hardware, such as reconstruction from sparse-arrays [37]. Handheld ultrasound devices are still limited to intermittent use due to excessive heating and low battery life [38] and come in a probe-like form factor, which limits continuous every day use. Future opportunities lie in improve the usability and wearability factors.

While there have been works using flexible electronics [39]–[41], these are still proof-of-concept. Doppler ultrasound devices cannot detect blood flow when the beam is perpendicular to the direction of blood flow. Imaging ultrasound does not have this drawback, but the beam must be directed toward the artery of interest. Future opportunities lie in miniaturizing systems into a wearable form factor and designing bio-compatible and skin-conformable transducers with configurable insonation angles and artery alignment.

D. Force-based modalities

Principles. Force-based modalities passively transduce mechanical pressure (applied force over an area). The predominant force-based modalities are based on resistive, capacitive, and piezoelectric effects. Resistive sensors, including strain gauges and piezoresistive sensors, consist of patterned structures supported on an insulating flexible backing. As the material deforms from arterial distensions, the

electrical resistance changes. Capacitive sensors use diaphragms and pressure cavities. As the diaphragms and pressure cavities change in volume, electrical capacitance changes. Piezoelectric sensors rely on the electric potential generated across a piezoelectric material when a force is applied [42]. Piezoelectric sensors used in Ballistography (BCG), which measures the ballistic forces (upward recoil from blood moving through the descending aorta) from the heart [43], [44].

Design choices. Designing force-based modalities requires considerations of drift, pressure sensitivity, response speed, dynamic range, temperate and humidity sensitivity, and hysteresis. High drift leads to inaccurate measurements when comparing values across time. High pressure sensitivity implies higher SNR. High response speed allows capturing higher frequency waveform features. High dynamic range allows the capture of a broader range of values. High temperature and humidity sensitivity lead to inaccurate measurements across different environmental conditions. High hysteresis leads to irreversible performance deterioration over time. Capacitive sensors tend to have low drift, have high response speed, are less sensitive to temperature and humidity changes, and have lower hysteresis, but have low pressure sensitivity [42], [45], [46]. Resistive sensors tend to have lower drift and moderate pressure sensitivity but are sensitive to temperature and humidity, have lower response speed, and have higher hysteresis [42], [45], [46]. Piezoelectric sensors tend to have high pressure sensitivity, have high response speed, are less sensitive to temperature and humidity changes, and have less hysteresis, but have high drift [42], [45], [46]. Capacitive, Resistive and Piezoelectric sensors can have broad dynamic ranges, depending on the material and structural designs [45], [46]. While capacitive and resistive sensors can support measuring static and dynamic loads due to small drift and have stable calibration, piezoelectric sensors can only measure dynamic loads due to high drift [47]. Support for static loads could be helpful in quantifying contact pressure.

Future opportunities. While force-based modalities are also skin tone agnostic and can measure pressure waveforms, they are the most susceptible to motion artifacts and temperature changes. Capacitive, resistive, and piezoelectric sensors are the predominant force-based modalities, but promising technologies also leverage optical, triboelectric, and magnetoelastic effects [42]. However, these new technologies are still proof-of-concept and are not yet scalable. Opportunities lie in developing scalable materials that jointly improve drift, pressure sensitivity, response speed, dynamic range, temperate and humidity sensitivity, and hysteresis characteristics.

E. Extending the capabilities of a single modality

The standard setup and measurement protocols for single modalities have inherent limitations in signal quality and the amount of information provided. Controlling contact pressure, controlling hydrostatic pressure, using multiple sensors of the same modality, and building multi-modal systems may make systems more robust. Controlling for contact pressure helps to “normalize” measurements within and, to some extent, between subjects for repeatable features. Multiple sensors of the same modality can be placed along the arterial tree to attain additional features. Combining different modalities allows us to leverage advantages and information from each while mitigating some of the noise. Flexible electronics may make systems more wearable. We review the extension of single modalities (Fig. 2.).

1) Contact Pressure Control:

Principles. While signals acquired with different contact pressures do not have visible artifacts, varying contact pressures distort waveform morphology, as pressure on the tissue may limit arterial distensions. Low contact pressure generally results in fatter and wider

pulses, while higher contact pressure results in more “spiky” pulses [48]. Pressure varies linearly around mean arterial pressure (MAP) and non-linearly away from MAP. Controlling for contact pressure is a way to help “normalize” measurements within and between subjects.

Design choices. The important design considerations for contact pressure control devices are optimal contact pressure, tightening mechanism, and measurement protocol. Optimal contact pressure is often defined as the contact pressure that results in the highest signal amplitude. Signals with the highest amplitudes can be attained by controlling contact pressure around MAP on a millisecond scale [49]–[53]. Other works have explored controlling contact pressure through finger-pressing [54]–[56]. While optimal contact pressure is a valuable point for more consistent measurements within and across subjects, it is unsuitable for continuous measurement because it is highly uncomfortable and occludes blood flow.

Future opportunities. While the effect of contact pressure has been investigated, identifying repeatable contact pressure targets and mechanisms to control and validate adequate contact pressure has not been fully explored. Investigating lower and repeatable contact pressure targets, while maybe not optimal with the highest signals, will result in more comfortable devices and can provide more robustness in the measurement. Developing tightening mechanisms for the device needs to be user-friendly and automated using an actuator, including, for example, a velcro wrist strap that allows fine-adjusting [57]. Developing protocols and algorithms to determine when the device is ready to take a measurement will be crucial for robust systems.

2) Hydrostatic Pressure Control:

Principles. It is well known that vertical arm displacements affect BP measurements due to changes in hydrostatic pressure [58], [59]. Higher hydrostatic pressure leads to overestimation of MAP [60], [61]. Variations in hydrostatic pressure also affect features [62]–[64] such as Pulse Transit Time [65] and Pulse Wave Velocity [66]. Therefore, standard BP measurement protocols require strict arm positioning to align the heart’s left ventricle to the transducer [60].

Design Choices. Hydrostatic pressure must be accounted for in the measurement protocol. The primary design consideration for hydrostatic pressure control is how to account for it during measurement. Most devices measure BP at repeatable positions across individuals. Some devices leverage hydrostatic pressure for calibration and measurement through arm raises [61], [65]–[69] or adjusting forearm angle [70] at pre-measured heights or using accelerometers. However, while physical movements provide an easy way to vary hydrostatic pressure, the variation in BP may not be explained by the hydrostatic pressure alone [71]. Incremental changes in BP are greater in hypertensive individuals [71].

Future Opportunities. Most devices that account for hydrostatic pressure require the user to interact actively with the device, such as by raising the arm. However, this is not suitable for continuous everyday monitoring. Future opportunities lie in integrating these hydrostatic pressure variations into BP measurements that are appropriate for everyday use.

3) Multiple sensors of the same modality:

Principles. Multiple sensors of the same modality with the same source frequencies can be placed along the arterial tree to attain additional features, such as Pulse Transit Time (PTT). PTT is the time delay (on the order of milliseconds) between two waveforms at a proximal and a distal point along the arterial tree. PTT depends on the distance between measurement sites, arterial compliance, arterial diameter, and blood density [72]. PTT generally decreases as systolic blood pressure increases due to decreased arterial compliance. Multiple sensors of the same modality with different source frequencies

Operating Principles	Contact pressure control	Hydrostatic pressure control	Multiple sensors of same modality	Multiple modalities	Flexible Sensors
	<p>Transmissive Reflective</p>		<p>PTT wrist-finger</p>		<p>Sempionatto <i>et al.</i>, 2021 [39]</p>
Characteristics	Normalize applied pressure on tissue across individuals	Normalize sensor measurement site relative to heart	Place sensors with same or different source frequencies at different site of measurements	Place different sensors on different sites of measurement	Use stretchable and bendable materials
	<ul style="list-style-type: none"> + Repeatable features at optimal contact pressure (mean arterial pressure, MAP) - Optimal contact pressure is uncomfortable to wear and occludes blood flow 	<ul style="list-style-type: none"> + control and account for contribution of hydrostatic pressure on BP measurements - patient must place measurement site at heart level or use extra sensors to measure displacement 	<ul style="list-style-type: none"> + Supplemental features like Pulse Transit Time (PTT) - Inconvenient to wear two sensors 	<ul style="list-style-type: none"> + Leverage features from multiple modalities to mitigate disadvantages of single modality - Inconvenient to wear two sensors 	<ul style="list-style-type: none"> + Long term comfort - Large accompanying electronics

Fig. 2. Extensions to single modalities for more robust systems. Controlling for contact pressure “normalizes” measurements between subjects for repeatable features across individuals. Controlling for hydrostatic pressure negates the effect of vertical arm displacements on BP estimation. Multiple sensors of the same modality can be placed along the arterial tree to attain additional features. Combining different modalities allows us to leverage advantages and information from each. Flexible electronics may make systems more wearable.

can be used to probe and separate different layer contributions under the skin [57], [73]–[76] and thereby affects PTT [76]–[79]. Lower frequencies probe deeper into the skin and travel further to the photodetector.

Design choices. Designing multiple sensor systems with the same modality requires considering the site of measurement, fiducial points, and source frequency. The choice of site is dependent on convenience for everyday wear. Commonly used sites of measurement include along the arm, along the neck, and between limbs. Increasing the arterial distance between measurement sites may allow for a smaller percentage error for PTT since the pulse must travel a longer distance from proximal to distal points, but it may be less convenient. Fiducial points for PTT must be reliable. Common fiducial points are pulse peak, pulse foot, or maximum slope. Leveraging multiple frequencies requires choosing the appropriate frequency to probe the desired layer (arteries, arterioles, or modulation of the capillary density) and remove signals from other layers that contribute greatly to the received signal. For light-based modalities, green wavelengths tend to penetrate the capillaries and arterioles, and red and NIR wavelength signals penetrate the arterioles and arteries [80].

Future opportunities. Multiple sensor systems with the same modality can help mitigate the limitations in single sensor systems but can also introduce additional limitations. PTT requires two devices placed on the body. Thus, the form factor requires careful design to ensure the contact pressures can still be made repeatable [81] and that the device is convenient for the end user. Opportunities lie in finding measurement sites that allow for unobtrusive form factors and developing devices that accommodate smaller separation distances. For light-based modalities, signal quality also varies based on skin tone and obesity levels [82], [83]. Opportunities lie in accounting for skin tone for better arterial pulse extraction and mitigating the effect of contact pressure on fiducial point detection.

4) Multiple modalities:

Principles. Combining different modalities allows us to leverage

advantages and information from each. PTT can be measured using multi-modal approaches such as IPG with PPG [84], [85]. Arterial diameter and PWV can be measured using PPG with ultrasound [86]. In addition to the modalities mentioned above, other modalities such as Electrocardiography (ECG), Seismocardiography (SCG), and Phonocardiography (PCG) can be used to provide more information for BP estimation. Unlike the core modalities, ECG, SCG, and PCG features alone do not correlate well with BP. However, ECG, SCG, and PCG features can still be used to increase the amount of information for BP estimation. Electrocardiography (ECG) measures the electrical activity of the heart. ECG and PPG can be used to measure Pulse Arterial Time (PAT), representing the time difference between the electrical depolarization of the heart’s ventricle (R peak in ECG) and a fiducial point in the PPG waveform (usually the peak, maximum slope, or foot) [87]–[89]. The difference between PTT and PAT is the Pre-Ejection Period (PEP), which represents the time delay between the electrical depolarization of the heart’s ventricle and the opening of the aortic valve [90]. While PAT is a convenient surrogate for PTT, it is inaccurate due to the considerable variation in PEP. Exercise, medication, and posture changes significantly affect PEP [88]. PAT and PTT models may only be sufficient to explain individual-level BP changes [91]. Seismocardiography (SCG) measures the blood and valve vibrations of the heart [92]. SCG and PPG can measure PAT [93]. SCG, ECG, and PPG can separate PEP from PAT to compute PTT [94], [95]. Phonocardiography (PCG) measures the sounds and murmurs produced by the heart during contraction. PCG and PPG can be used to measure PTT [96], [97].

Design choices. Designing multi-modal systems requires consideration of the advantages and disadvantages of each modality/multi-modal system in terms of robustness and wearability. As noted above, while light-based systems are cheap and ubiquitous, they are affected by skin tone and obesity. While impedance-based systems are skin tone agnostic, they introduce electrical-specific limitations, such as requiring current injection, skin-electrode impedance matching, and

bio-compatibility. While sound-based systems allow blood velocity measurement and imaging, they incur high costs, have large form factors, and require careful considerations of insonation angle, artery alignment, and impedance matching. While motion-based sensors can measure pressure waveform, they suffer from temperature variations and are highly susceptible to motion artifacts. While other modalities can provide extra features that may correlate with BP, the added value of additional modalities may or may not outweigh the decrease in wearability.

Future opportunities. Choosing modalities with less attribute overlap will improve robustness, as they can compensate for the other's shortcomings. However, the contribution of each modality to robustness and the reasons are not entirely clear. Opportunities lie in performing calibration studies to determine which features and modalities are essential and can enhance the measurement's overall robustness.

5) Flexible electronics:

Principles. Motion artifacts and mechanical mismatches between rigid devices and skin lead to measurement errors. Flexible electronics offer a potential solution by using stretchable and bendable materials that conform to the skin and are comfortable for long-term wear [33], [39], [41], [98]. These devices consist of substrates, interconnects, and active sensing components [99]. Substrate materials are made of polymer films and stretchable elastomers. Interconnects comprise solid metals (thin ribbons and filaments), conductive polymers, carbon, and conductive liquid metals. The materials of active sensing components are modality-dependent.

Design choices. Designing flexible sensors requires considering flexibility, adhesion, gas permeability, and breathability. Adhesive materials help maintain sensor-skin contact and reduce some motion artifacts, but high adhesion may lead to adverse skin reactions. Gas-permeable and breathable materials help mitigate inflammation and irritation from blocked sweat glands. For a more detailed exploration into flexible sensor design choices, we refer the reader to [100].

Future Opportunities. Although flexible electronic patches tend to be small, they can require large accompanying electronics and power sources. Many patches use flexible cables and pins to connect to external electronics [100]. Opportunities lie in miniaturizing the bulky external electronics. In addition, sliding or detachment of devices results in signal deterioration during long-term use. While adhesion is a solution, this comes at the cost of skin irritation, which itself would limit long-term use. Opportunities lie in identifying and developing both bio-compatible and adhesive materials.

III. PRE-PROCESSING AND FEATURE EXTRACTION

Pre-processing ensures adequate signal quality and improves feature saliency for downstream tasks, achieved using signal processing filters and data selection criteria. Feature extraction attains low dimensional representations of signals for better BP estimation and is achieved through arterial waveform analysis and pulse wave decomposition. Volume, velocity, and pressure waveforms can leverage similar pre-processing and feature extraction techniques because they share similar morphologies but with nuanced differences. In short, volume over time is the product of arterial area and velocity, and pressure is related to velocity through the transmission line model [101, Ch. 9]. Velocity waveforms comprise forward pulses superimposed with inverted reflected pulses [101, Ch. 9]. Pressure and volume waveforms comprise forward pulses superimposed with non-inverted reflected pulses [101, Ch. 9]. We review pre-processing and feature extraction techniques for arterial waveforms.

A. Pre-processing

Principles. The general framework for pre-processing is to determine sources of noise, apply signal processing and filtering techniques, quantify or qualify signal quality, and remove signals that do not meet pre-specified criteria (Fig. 3). Sources of noise originate from hardware (high-frequency powerline interference), measurement (motion, ambient light, missing data, and sensor saturation), and physiology (respiration rate, background absorbance from different tissue types, and venous pulsations). Signal quality can be determined visually or quantitatively by signal quality indices (SQIs), which consist of temporal, frequency, and statistical quantities. Generally, "expert" annotators have visually classified signal quality based on feature saliency. However, there is no consensus on signal quality guidelines [102]. For SQIs, a threshold or range can be defined to classify good and bad signals. The bad signals are usually discarded.

Implementation. Designing data and signal processing filters requires understanding whether the underlying physiological signal can be recovered, the frequency band that the noise manifests in, and attributes that constitute a "good" signal. Frequency filters (i.e., low-pass, band-pass, band-stop, high-pass) can remove certain hardware and physiological noise. For example, one form of hardware noise includes high-frequency powerline interference, which can be mitigated by high-pass filtering [103] while another form, like 60Hz line noise, can be removed with a band stop filter. Physiological noise (respiratory rate, background absorbance and scatter from different tissue types, and venous pulsations) can be somewhat mitigated using high-pass and low-pass filters [21]. However, frequency filters cannot remove noise that includes the frequency band in the physiological signal of interest. Measurement noise (acquisition artifacts from motion, ambient light, missing data, and sensor saturation) generally cannot be recovered and is usually discarded. Such corrupted signals contain minimal information about underlying physiology and can be identified visually or through temporal, frequency, and statistical quantities [21]. Signal processing filters can also be used to increase feature saliency. The 4th-order Chebyshev-II filter has improved the saliency of various systolic and diastolic features in PPG signals [104]. Feature saliencies are used to determine signal quality and develop SQIs. The Skewness index has been found to outperform other statistical features for signal quality [105]. Template matching using dynamic time warping [106], spectral quality [107], and goodness metric [108] have also been used to determine signal quality. Support vector machines [109], hierarchical classifiers [110], and convolutional neural networks [111], [112] have been used to determine signal quality based on extracted or learned features.

Future opportunities. Some works investigated how signal processing filters improve BP estimation [113], [114]. However, the assumption that high saliency results in more robust estimations has not been explored in depth. Opportunities lie in optimizing signal processing filters for better BP estimation. Biology, physiology, and external factors affect waveform morphology. Using metrics based on feature saliency may exclude particular demographics and harm model generalizability. Some features like the dicrotic notch may not be visible in elderly subjects [115]. Opportunities lie in finding SQIs that are equally effective for all demographics.

B. Feature Extraction

Features are low-dimensional representations of waveforms. The goal is to extract features that will be helpful for BP estimation. It is important to understand how arterial waveforms manifest to determine what features are useful. We describe the intuition of the waveform as the superposition of forward and reflected waves, key parameters in the waveform, and how physiology changes the

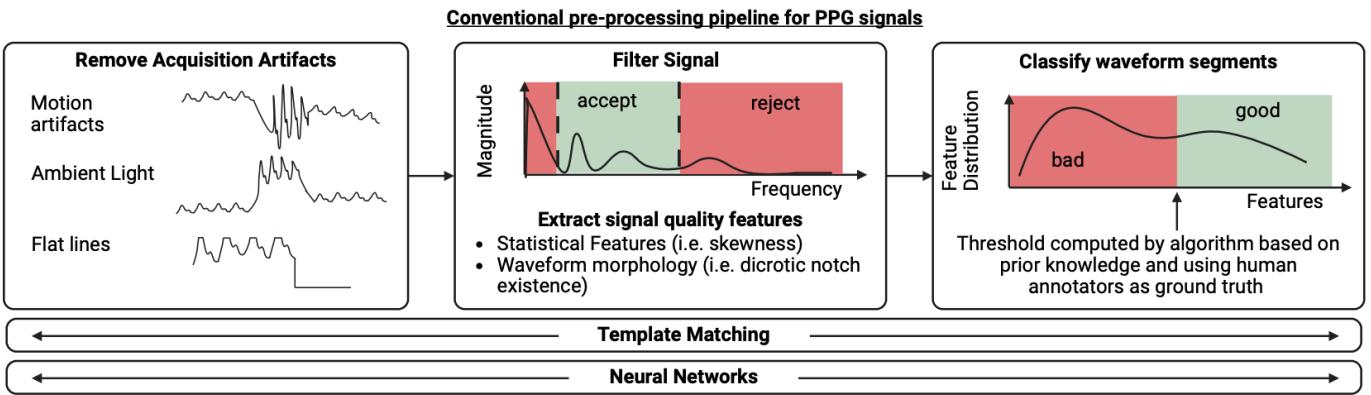


Fig. 3. Steps to pre-process PPG data. Pre-processing involves designing algorithms to remove artifacts, filter signals, extract signal quality features, and classify waveform segments. Artifacts include motion artifacts from the device sliding around the wrist/finger, ambient light from surroundings that seep into crevices along the device-skin interface (for light-based modalities), and flat lines from missing data or exceeding the device acquisition range. Then, filter signals and extract signal quality features. These can be done interchangeably. Filtering methods could include using a band-pass filter to remove unwanted frequencies. Signal quality features can determine whether a signal is suitable for BP estimation. These include statistical features and fiducial point detection. Then, the waveforms are classified in a supervised manner by computing a threshold based on prior knowledge and using human annotators as ground truth. Finally, template matching and neural network-based methods have been developed to remove artifacts, filter signals, extract features, and classify signals in an end-to-end manner.

waveform morphology. Then, we delve into methods to extract DBP and SBP features.

Principles. During each heartbeat cycle, blood is ejected from the left ventricle into the aorta, creating an arterial pulse pressure wave. As this pressure wave travels down the vascular network, it reaches junctures that result in major reflections between the thoracic and abdominal aorta (R_1), and abdominal aorta and common iliac arteries (R_2). The dicrotic notch appears between the first and second major reflections (Fig. 4.A). Other reflections from minor junctures and re-reflections ($\{R_3, R_4, \dots, R_n\}$) also contribute to the waveform morphology [116] (Fig. 4.A). The resulting arterial waveform consists of several key parameters. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) are the maximum and minimum of the waveform during each cardiac cycle. The difference between SBP and DBP is pulse pressure (PP). MAP is the area under the arterial waveform curve. While the exact physiological mechanisms are complex and depend on physiology, measurement site, and measurement conditions, waveform morphology and MAP generally change based on vascular tone (VT), stroke volume (SV), systemic vascular resistance (SVR), heart rate (HR), cardiac output (CO), and compliance (C). VT is the degree of constriction (vasoconstriction) or dilation (vasodilation) of the smooth muscle in the vascular walls [117]. SVR is the resistance to blood flow by the systemic circulation, or the zero-frequency component of vascular impedance [118]. SVR affects wave reflections, which is in turn affected by the mismatch between input impedance and terminal impedance. SVR can also affect MAP, which alters the spectral components of impedance due to pressure-dependent arterial stiffness. Vasoconstriction results in higher SVR and vasodilation results in lower SVR [117]. SV is the amount of blood ejected by the heart's left ventricle in a single contraction during systole. HR is the number of times the heart beats per minute. CO is the volume of blood that the heart pumps out in one minute and is the product of SV and HR, $CO = SV \times HR$. C expresses the ability of an artery to distend with increasing transmural pressure [119]. The lower the C, the more stiff the arteries. Pulse pressure (PP) is the ratio between stroke volume (SV) and compliance (C), $PP = \frac{SV}{C}$. MAP is the product of SVR and CO, $MAP = SVR \times CO$.

Feature Extraction Methods. Features can be extracted using pulse wave decomposition and arterial waveform analysis. In pulse wave

decomposition, multiple Gaussian curves, representing the forward and reflected waves, are superimposed to fit the waveform [120]–[122]. Hyperbolic secant functions can be used and may provide better and more stable features because they represent possible solutions to the Moens-Korteweg equation [123]. In waveform analysis, features are extracted by finding fiducial points using peak detection algorithms and the derivatives of the signal [124]–[126].

DBP Features. Typically, DBP is determined by VT and HR [117], [119] (Fig. 4.B1). Vasodilation (lower SVR) results in faster decay (smaller decay constant) in the diastolic phase [117], [127], [128]. Vasodilation also increases C, as the arteries can accommodate changes in blood volume. Intuitively, PWV decreases as C increases, the time difference between reflected waves is larger, and the resulting decay constant is smaller. However, an increase in SVR can increase PWV due to stiffer arteries only if MAP increases. In some cases, there would be no increase in PWV if the increase in SVR is associated with reduced CO with no change in MAP. The minimum point in the cardiac cycle (DBP) is determined by when the next heartbeat occurs [117]. The earlier the heartbeat occurs, the higher the DBP [117]. In practice, the SNR of the decay constant over small time scales could be significant, and the decay constant needs to be estimated over longer time scales of minutes due to confounding wave reflections [129]. HR must also be determined over longer time scales (seconds) for reliable measurements.

PP and SBP Features. PP is determined by SV and C [117], [119]. Since $SBP = DBP + PP$, SBP is therefore determined by SV, C, and DBP. For a fixed C and DBP, higher SV leads to higher PP and SBP. Intuitively, SV affects waveform scaling (Fig. 4.B2). Increasing SV vertically stretches the waveform [119], [130]. The vertical stretch of the waveform affects the difference between the maximum and minimum value of the waveform (systolic amplitude, SA), the time interval from the foot of the waveform to the peak of the forward wave (systolic upstroke time, ST), and the slope at the slope during the systolic upstroke (systolic upstroke slope, SUS) [131]. SA correlates with SBP [132]. ST correlates with changes in SBP and improves the accuracy of PAT systems [131], [133]. A smaller ST generally leads to a larger SUS slope for the same SV [131]. For a fixed SV and DBP, a smaller C leads to higher PP and higher SBP (Fig. 4.B2). Intuitively, smaller C increases PWV as predicted by the Moens-Korteweg equation, leading to smaller time intervals between

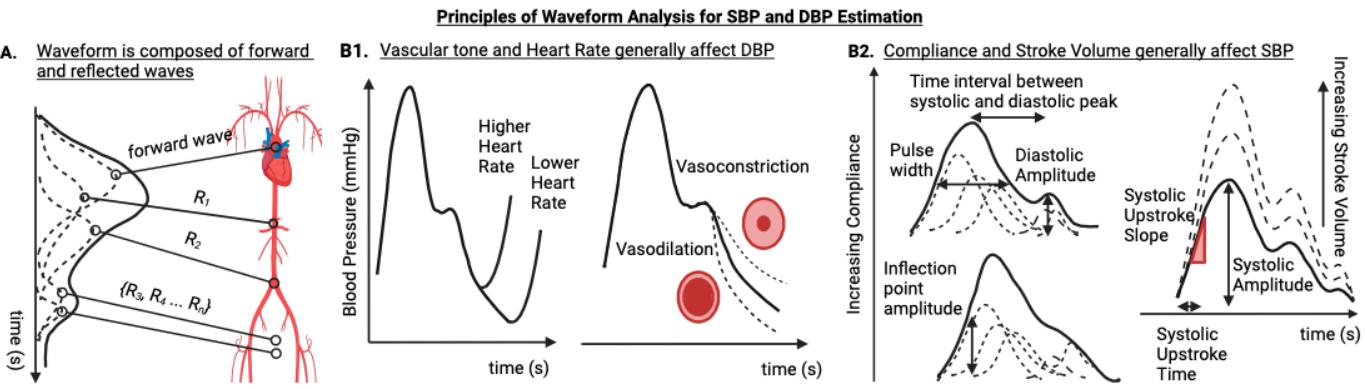


Fig. 4. Principles of Waveform Analysis for SBP and DBP estimation. A. The arterial waveform is the superposition of forward and reflected waves from junctures throughout the arterial tree. R_1 represents the major reflection from the juncture between the thoracic and abdominal aorta. R_2 represents the major reflection from the juncture between the abdominal aorta and common iliac arteries. $\{R_3, R_4, \dots, R_n\}$ represents the set of minor reflections from other junctures that contribute to the change in waveform morphology. B1. Vascular tone and heart rate generally affect DBP. Vascular tone determines the steepness of the diastolic decay, and heart rate determines the termination of the diastolic decay. B2. Compliance and stroke volume generally affect SBP. Increasing compliance decreases pulse wave velocity, leading to larger time intervals between forward and reflected waves. Both affect waveform morphology.

forward and reflected waves and higher wave superpositions. C affects the pulse width of the waveform, the ratio between the person's height (h) and the time interval between the systolic peak and the diastolic peak ($\Delta t_{SP \rightarrow DP}$) (stiffness index, $SIx = \Delta t_{SP \rightarrow DP}$), the ratio of the difference between SA and amplitude at the inflection point (IA) and SA (augmentation index, $AIx = \frac{SA - IA}{SA}$) [101], [124, Ch. 9]. Pulse width of the systolic phase correlates with SBP [134]. For the same person's height, SIx correlates with PWV [135]. AIx correlates with age, a surrogate for C [101], [124].

Future opportunities. Arterial waveform analysis can work well on signals with apparent morphological features. However, in many cases, not all features can be identified accurately. Alternatives to arterial waveform analysis exist, such as using the number of crossings between the waveform and arbitrary lines [136]. Pulse wave decomposition could offer a better solution. However, it requires long compute times for fitting and may be unreliable when the decomposed pulses are close together. Opportunities lie in improving the speed and accuracy of existing pulse wave decomposition techniques and developing new methods to analyze waveform formation. Some features have been investigated on an individual level, whereas others have been investigated between individuals. Some features are more repeatable (multiple measurements in identical conditions for the same subject), reproducible (multiple measurements in different conditions for the same subject), and reliable (multiple measurements in different conditions for different subjects) [115], [137]. Features more indicative of C for large arteries were more repeatable than those strongly influenced by smaller arteries, possibly due to the susceptibility of motion artifact and small changes in small arteries [115]. [115] found PWV was the most reproducible long-term. Opportunities lie in understanding what features can be used to generalize reliably within and across individuals and contexts. While VT, HR, SV, and C are interrelated through BP regulation mechanisms, experimental designs such as exercise, use of drugs, temperature exposure, and diversity in ages or diseases (i.e., hypertension, atherosclerosis) could change how physiological factors interact and may disproportionately affect each feature's performance. Features that estimate BP well in one experiment may not translate to another. Opportunities lie in designing experiments that isolate each component's contribution to estimation and characterizing their usefulness for different measurement conditions. Opportunities also lie in developing an understanding of feature variations within and

across individuals on medication, such as hypertension drugs as a function of the drug type, drug dose, and the timing before and after drug dosing.

IV. ESTIMATION ALGORITHMS

Estimation algorithms take manually extracted features or learn features from waveforms to predict BP in a supervised setting. The three main classes of algorithms are classical machine learning (classical ML), physiological, and deep learning models. We review the three classes of algorithms (Fig. 5).

Principles. Classical ML algorithms use manually extracted features. Examples include linear regression (LR) [133], [138]–[143], support vector regression (SVR) [143]–[145], random forests (RFs) [141], [146], [147], adaptive boosting (AdaBoost) [33]. Deep Learning algorithms leverage the high capacity of neural networks to learn relationships from manually extracted features or learned low-dimensional representations of waveforms. Examples include multilayer perceptrons (MLPs) [143], convolutional neural networks (CNNs) [32], [148], [149], long-term-short-term networks (LSTMs) [57], CNNs with LSTMs [150], [151], and transformers [152]. We refer the readers to standard introductory machine learning texts to learn the fundamental concepts [153], [154]. Physiological models incorporate physiological domain knowledge into algorithm design (Fig. 5.B). Examples include lumped parameter (i.e. Windkessel) [155]–[157], transmission lines [158], and PWV [159] based on the Moens-Korteweg equations [101, Ch. 3]. Lumped parameter models represent cardiovascular parameters, volumetric flow, and pressure as circuit elements (capacitors, resistors, and inductors), current, and voltage. Transmission line models distribute resistance, capacitance, and inductance continuously throughout the circuit in infinitesimally small elements. Transmission line models allow us to individually model the forward and reflected waves at arterial junctions using the pulse wave decomposition concept [101, Ch. 9]. Lumped parameter and transmission line models are used to derive closed-form solutions for SBP and DBP. Features from sensor data are used to estimate the unknown parameters. The estimated parameters are used to obtain estimates of SBP and DBP using the closed-form solutions. The Moens-Korteweg equation models the relationship between pulse wave velocity (PWV) and arterial properties (elastic modulus, wall thickness, vessel radius, and blood density). PWV can also be calculated by dividing the distance by time required for the arterial

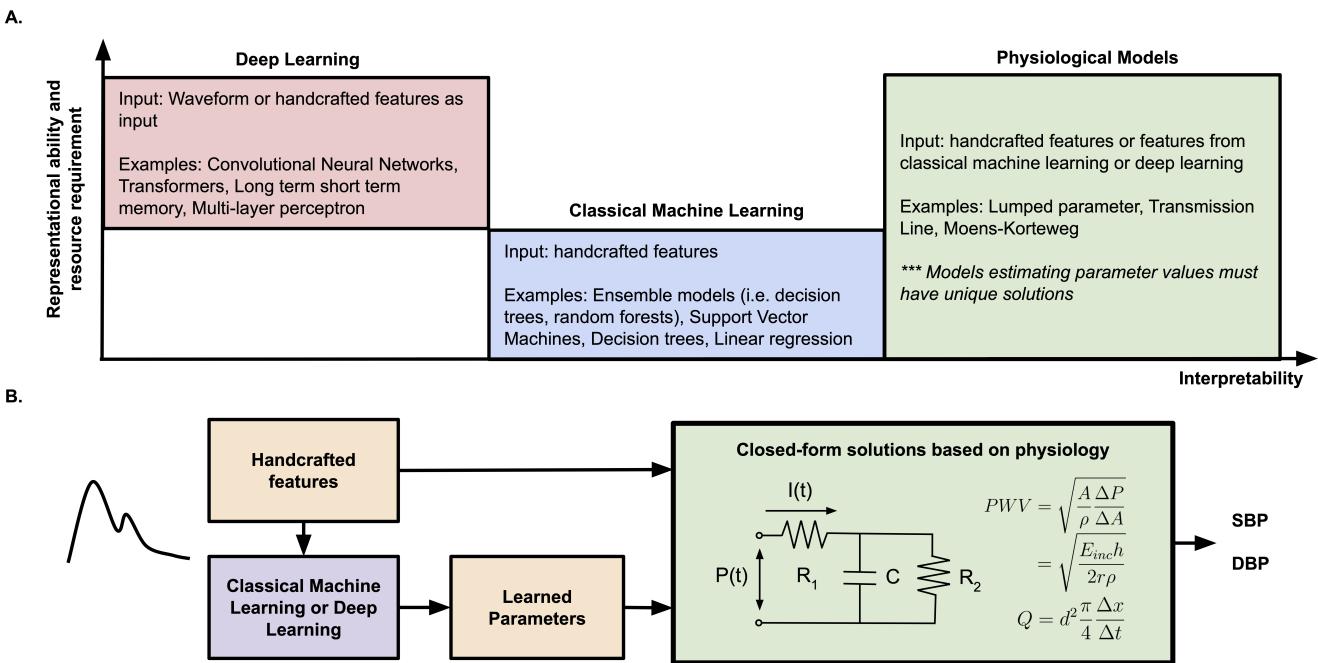


Fig. 5. *Types of estimation algorithms.* Classical machine learning (classical ML) algorithms use manually extracted features. Deep learning algorithms leverage the high capacity of neural networks to learn relationships from manually extracted features or learned low-dimensional representations of waveforms. Physiological models use features from classical ML and deep learning and incorporate physiological domain knowledge into algorithm design. A. Representational ability (how complex a pattern can be learned) or resource requirement (computational and data necessary to train and test) versus interpretability for Classical ML algorithms, deep learning algorithms, and physiological models. B. Physiological model pipeline - handcrafted features or learned features are used to predict parameters from a physiological model.

pulse to travel between a distal and proximal point on the arterial tree.

Design Choices. While a wide array of machine learning models can be used, several factors should be considered when designing estimation algorithms, namely representational ability, resource requirement, and interpretability (Fig. 5). Representational ability indicates the complexity of patterns and relationships the model can express. Highly non-linear relationships between features and BP require models with higher representational ability. Resource requirements indicate the computing power and data needed to train and deploy the model. Higher resource requirement may potentially increase deployment costs and decrease access. Interpretability is understanding the inner mechanics of why and how the model generates predictions. Interpretable models help identify whether BP predictions are realistic and may potentially help with prognosis for downstream applications.

Classical ML approaches have moderate representational ability and require careful feature engineering but require fewer resources and can be interpretable. LR has less representational ability because it is only limited to linear relationships, but the coefficients are directly interpretable and confidence intervals can be computed for hypothesis testing. SVR has a higher representational ability than linear regression by allowing for non-linear kernels in the decision function and can be effective when there are more features than the number of samples, but it is a black box. DTs are easily interpretable since each tree outputs a floating point number, but will decrease interpretability as the model becomes more complex. RFs have higher representational ability than SVR but tend to be uninterpretable because they can consist of hundreds or thousands of DTs. AdaBoost has a higher representational ability than RFs but suffers the same drawbacks in interpretability as RFs. Deep learning approaches have high representational ability and can automatically

learn features from waveforms but are generally not interpretable by design. Like classical ML, MLPs can learn non-linear relationships between input features and BP. On the other hand, CNNs, LSTMs, and transformers learn from waveforms. CNNs capture local morphological features. LSTMs and transformers capture time-dependent features. Combining CNNs and LSTMs allows the extraction of both spatial and temporal data. Transformers can capture both local and global features due to the attention mechanism. Transformers also avoid recursion in LSTMs due to the attention mechanism by processing the input as a whole. Physiological models inherit the advantages and disadvantages of classical ML and deep learning approaches and could yield meaningful physiological parameters. For some models (i.e. lumped-parameter, transmission-line models), there may be an infinite number of arterial properties and topology parameters that can fit a model well [160], [161]. Careful system and algorithm design ensures well-posed problems and uniqueness of parameters. The generalizability is constrained by the degree to which the model accurately represents human physiology and blood flow, especially in diseased subjects. Transmission line models tend to be more accurate and have higher representational ability than lumped parameter models. The Moens-Korteweg equation provides a direct relationship between the arterial elastic modulus and a measurable quantity (PWV). However, the generalizability within and across individuals is questionable due to the variance in wall thickness, vessel radius, and blood density, especially if the individual is on hypertension medication, which may cause these variables to change with the dose and time when the medication is taken.

Future opportunities. Extensions to improve generalizability such as transfer learning, have been explored [57]. However, which datasets are useful remains an open question. Opportunities lie in quantifying the mutual information between waveforms to determine effective data augmentation datasets. Estimation algorithms

should have balanced errors across covariates (i.e., gender, age, blood pressure range). Some works have identified algorithmic and dataset bias concerning blood pressure range [162]. Opportunities lie in understanding the origins of these biases and correcting them using techniques such as inverse probability weighting [163] or data augmentation [164]. A wide range of segmentation window lengths is used for feature extraction or representational learning. Although many works empirically determine the right segmentation window length for the best accuracy, minimal work is devoted to understanding why these lengths are optimal. Opportunities lie in understanding how and why accuracy varies for different segmentation lengths. While physiological models can represent cardiovascular disease parameters, solutions to the physiological models can be non-unique if the problem is ill-posed. Opportunities lie in validating that these parameters are physiologically accurate with reference measurements.

V. EVALUATION

While the rapid volume of research efforts is encouraging, there has somewhat been lacking standards and expectations in study designs and experimental evaluation. This makes it extremely difficult for the community to parse out the most effective and important emerging ideas. We perform a systematic review and provide insights into evaluating device accuracy and study design based on the standards specified by ANSI/AAMI/ISO 81060-2:2019 [60] (See I). Despite well-defined standards, large heterogeneity exists in datasets, devices, validation protocols, and algorithms. The objective is to demonstrate that calibration technique and BP distribution are confounders in BP estimation error and to account for them using a metric we call “explained deviation” (which is closely related to coefficient of determination, R^2 , an oft-used statistical measure of quality of a fit). We describe validation standards, calibration techniques, accuracy measures, explained deviation, an open-contribution benchmark, and results.

Validation Standards. Device approvals are based on meeting the requirements specified by some validation standard such as from the American National Standards Institute/Association for the Advancement of Medical Instrumentation/International Organization for Standardization (ANSI/AAMI/ISO) [60], British Hypertension Society (BHS) [165], Institute of Electrical and Electronics Engineers (IEEE) [166], and European Society of Hypertension (ESH) [167], [168]. These standards include details about maximum allowable error, data distribution, test protocols, and reference devices. There have also been efforts to synthesize the standards [169]. While each standard specifies a unique combination of requirements, the common goal is to provide a minimum requirement for an accurate device. Devices requiring cuff initialization warrant different evaluation protocols because an individual’s initial BP is already known [14], [166], [169], [170] specify protocols and requirements for cuff initialized devices.

Calibration technique. We identified that there are three commonly used calibration methodologies: (a) subject split, (b) personalization, and (c) record split without personalization. Subject split involves using different subjects for the training and testing sets. A representative example is if 100 subjects are in the dataset, 50 subjects are used for training and the other 50 subjects for testing. Personalization allows subject data (a record) in training and testing sets but maintains causality in the data – training data only consists of data acquired before test data points. In other words, the training data is selected as the first portion of the record, and the testing data is selected as the latter portion, often with a clear, well-defined gap between training and test data. The training and testing data certainly do not overlap. For example, if a record for a single subject is one month long,

data from the first day is used for training, and the rest is used for testing. This is used when the model’s generalization ability across the population of interest is unsatisfactory, so an individual-level model is fitted and recalibrated at arbitrary intervals due to physiological changes that may render the model inaccurate. Finally, some studies do not fall into these two categories, involving overlapping or non-causally split data in the training and testing set from the same subjects, thereby introducing data leakage. Examples are randomly sampled segments from a subject in training and test sets and fitting a model to all data to find correlations. We call this practice as record split without personalization. This type of calibration is not realizable in real-world practice because it violates causal constraints. Therefore, these studies were excluded from our analysis.

Accuracy measures. Approval of BP devices requires meeting two criteria: *Criterion 1* - the differences in individual paired determinations of the device under investigation and the mean observer references should have a mean difference of $\leq 5\text{mmHg}$ and a standard deviation of $\leq 8\text{mmHg}$. *Criterion 2* - The maximum permissible standard deviation of the difference between averaged readings from the device under investigation for each subject and observer references should be less than a particular value specified in ANSI/AAMI/ISO 81060-2:2019 Section 5.2.4.1.2 [60]. For instance, a mean difference of 0mmHg should have a standard deviation value of less than 6.95mmHg. To simplify the analysis, we adopt Criterion 1 because not all studies may follow the multiple measurement protocol.

Explained Deviation. BP distribution in the dataset is a primary source of heterogeneity, making it challenging to make reliable quantitative comparisons across datasets, devices, and algorithms. For example, it is clear that if the experimental studies were performed on a homogenous population without much BP variation, then predicting BP for this population would be easier. In contrast, when studies recruit a truly representative population with typically much larger BP variations, predicting BP on these datasets is a lot more challenging. Consequently, the desire to meet or beat state-of-the-art metrics may encourage collecting non-representative, somewhat homogeneous subjects with lower inter-subject BP variations. One way to explicitly account for these differences in BP distributions across datasets is to normalize the measured BP errors by the standard deviation (STD) of BP in the dataset. This statistic, which we call “explained deviation” (ED) (Fig. 7) is related to the commonly used “R squared” (R^2) metric in statistics through the relationship $ED = \frac{1}{\sqrt{1-R^2}}$. When calibration is performed using subject split, $ED = \frac{\sigma_{pop}}{\sigma_{pop,\epsilon}}$ where σ_{pop} is the STD of BP measured with a gold standard and $\sigma_{pop,\epsilon}$ is the STD of BP prediction errors. When calibration is performed using personalization, one needs to account for the fact that information about that particular subject’s BP is known – therefore in this case: $ED = \frac{\sigma_\Delta}{\sigma_{\Delta,\epsilon}}$, where σ_Δ is the STD of BP changes and $\sigma_{\Delta,\epsilon}$ is the STD of BP change prediction errors. It is well understood that in studies that perform personalization, the longer the gap between training data and test data, the harder the BP estimation problem becomes [171]. As a result, a popular surrogate to keep track of when evaluating personalization-based techniques is the time between calibration and test (Δt), and maximum calibration periods for acceptable error limits have been explored [171]. However, large Δt does not necessarily lead to large BP changes, and the rate of physiological change may be individually and temporally dependent. In general, the higher the ED, the better the device performs. An ED of 1 indicates that the estimator performs no better than one that predicts a constant value at the mean for subject split studies and at the initial calibration value for personalization studies. ED determines how much the device and algorithm “explains” the BP

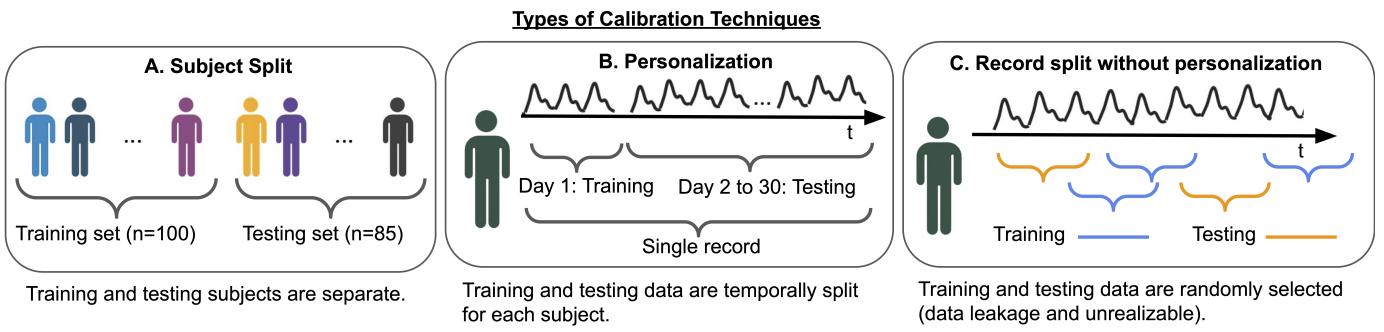


Fig. 6. Three different Calibration Techniques are present in Wearable BP literature. A. Subject split - training and testing subjects are mutually exclusive. B. Personalization - for each subject's record, the training set always appears before the testing set. C. Record split without personalization - for each subject's record, the training set and testing set are not mutually exclusive (i.e., overlapping windows), or the training set does not always appear before the testing set, which is not realizable.

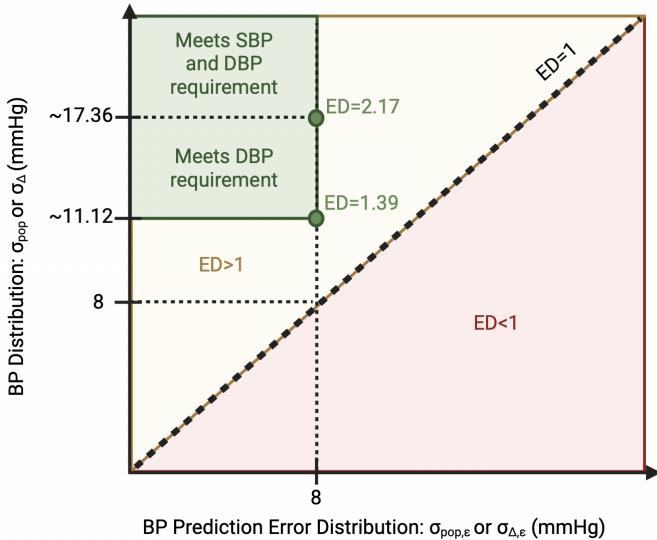


Fig. 7. ED determines how much the device and algorithm “explains” the BP distribution. Explained Deviation (ED) normalizes error by taking the standard deviation (STD) of BP into account. For subject split, $ED = \frac{\sigma_{pop}}{\sigma_{pop,e}}$, where σ_{pop} is the STD of BP in the dataset measured with a gold standard and $\sigma_{pop,e}$ is the STD of BP prediction error. For personalization, $ED = \frac{\sigma_{\Delta}}{\sigma_{\Delta,e}}$, where σ_{Δ} is the STD of BP changes and $\sigma_{\Delta,e}$ is STD of BP change prediction errors. The x-axis represents the respective denominator STD in each of the above calibration methodologies, while the y-axis represents the numerator STD. We approximated the minimum STD of SBP and DBP required to meet the standard as 17.36mmHg and 11.12mmHg, corresponding to an ED of 2.17 and 1.39 for an error of $\pm 8\text{mmHg}$.

distribution over and above a naive algorithm that predicts the mean BP as the estimate. We can compute confidence intervals and perform hypothesis testing (See II). Furthermore, we derive relationships between the minimum ED required to meet the standard - 2.17 for SBP and 1.39 for DBP (See III).

A. Results and Discussion

We performed a systematic review (Fig. 8) and evaluated studies using our adopted statistic. We identified 2510 articles from our database search and 29 articles from other sources. After adjusting for duplicates by code, 2180 remained. We inspected the full articles and excluded 2008 articles based on the inclusion criteria. Likewise, we included 173 unique articles and 67 additional studies (from articles that report multiple studies). From the 239 studies, we determined

that 91 studies performed subject split and 148 studies performed personalization. Ultimately, we included 59 subject split studies that reported STD of BP and 11 personalization studies that reported STD of BP changes. We observed that 42% of the studies performed record splits without personalization and, therefore, suffered from data leakage. These studies reported remarkably low errors. However, record splits without personalization do not obey causal constraints and are impractical. This practice may have resulted from publication bias, in which only the studies with good results are published [172], [173]. Moving forward as a community, we must avoid performing studies with record splits without personalization since the results of these studies tend to be much better while leaving no means to replicate them in the real world. We observed that only 27% of the studies reported sufficient power. Of subject split studies, 65% reported STD of BP in the dataset compared to 7% of personalization studies that reported STD of BP change in the dataset. We provide a complete list of included studies with extracted parameters and excluded studies in <https://wearablebp.github.io/bib>.

Distribution of Study Parameters. We explored the distribution of different study parameters for each calibration technique (Fig. 9). We observed that the most common sensor data configurations used for BP estimation are Photoplethysmography (PPG) and Electrocardiography with Photoplethysmography (ECG+PPG). For datasets, we observed that an overwhelming number of studies used internal databases, and the most common publicly available database was Medical Information Mart for Intensive Care (MIMIC) [174]. For algorithms, we observed that the most common class is classical ML. For study characteristics, we observed that most subject split studies were observational. In contrast, personalization studies were split evenly between observational and interventional studies. For the number of testing subjects, most datasets contain less than 300 testing subjects, and the difference between subject split and personalization studies is insignificant. For study power, the difference between subject split and personalization studies is insignificant. For STDs of SBP and DBP, subject split studies tend to have larger STDs of BP compared to STDs of BP changes in personalization studies (SBP: $p=2.2e-2$ and DBP: $p=1.0e-2$). For SBP and DBP EDs, subject split studies tend to have higher ED than personalization (SBP: $p=4.3e-6$ and DBP: $p=1.7e-4$).

Benchmarking. To tackle the large heterogeneity in study design, we introduce and referee the first open-contribution BP estimation benchmark available at <https://wearablebp.github.io/benchmarks>. We aim for valid (produce statistically significant results across different datasets), reproducible (provide code, data, experimental protocols, and device details), and accurate (high ED) results. To start with, we selected four algorithms that report high

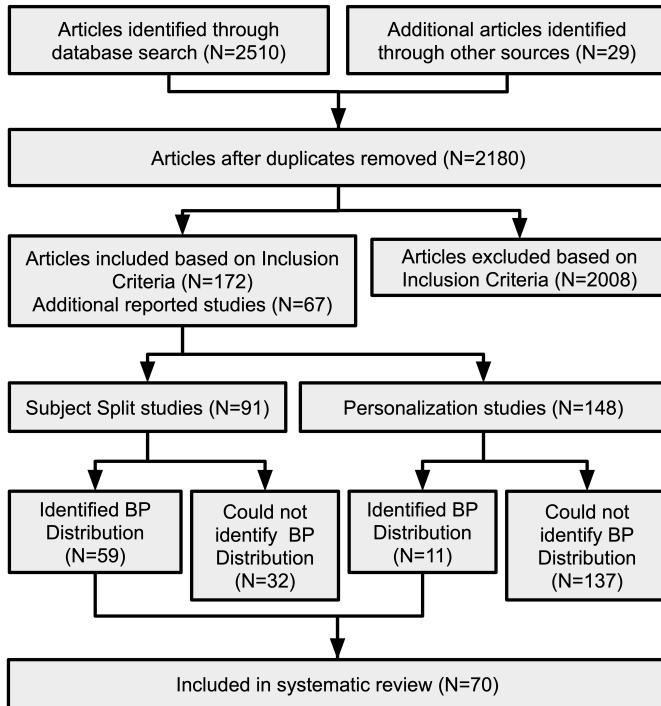


Fig. 8. Flow chart for systematic review. Based on the inclusion criteria, we included 70 studies in our systematic review (59 subject split studies and 11 personalization studies). N indicates the number of records, and n indicates the number of studies. Some articles report multiple studies and are reported as additional studies.

ED, namely algorithms presented in [146], [114], and [151] for benchmarking on the MIMIC, PPG-BP, and VitalDB for PPG (and ECG) sensor data. We chose public datasets with sufficient power, namely MIMIC [174], PPG-BP [175], and VitalDB [176]. We base our implementation on code provided on GitHub (for [146], [114]) and the details of the algorithms presented in the paper (for [143] and [151]). We implement the algorithms to the best of our ability and display the results in Appendix IV. We provide an opportunity to contribute to this benchmark and a codebase for feature extraction machine learning pipelines, deep learning pipelines, and data visualization scripts to allow for more transparent reporting and streamlined testing available at <https://github.com/wearablebp/wearablebp-benchmarks>. Currently, benchmarks only cover PPG and ECG+PPG, as there are limited publicly available databases for other sensors. However, with the rapid pace of research activity in this area, we expect that to change. As other wearable systems become more standardized and datasets are publicized, the benchmark will be expanded to include them.

BP Distribution vs Error Distribution. To determine whether the BP distribution in a dataset affects the reported errors in that dataset, we regress the STD of the reported errors on the STD of BP for subject split and STD of BP change for personalization together (Fig. 10). We removed one outlier based on Cook's distance (z-score greater than 4). The slope coefficients are significant (SBP: $p=8.6e-5$ and DBP: $p=1.9e-2$) and account for 25-35% of the reported error values. The results indicate that lower error is harder to achieve for larger BP distributions and confirm that BP distribution is a significant confounding factor in comparing the accuracy of methodologies across datasets. Indeed, several studies show the STD of the error is nearly as large or larger than the standard deviation of the BP distribution, which indicates that these algorithms are performing no better than a naive estimator that provides the dataset mean BP as

the estimate independent of sensor data collected. Therefore, we must take BP distribution into account when comparing studies.

Time between calibration and test vs Error Distribution. We investigated the accuracy of the personalization changes over time across all observational studies by regressing the error versus the time between calibrating and testing the model (Δt) on a logarithmic scale ($\log_{10}(\text{seconds})$) (Fig. 11). We include all observational personalization studies that reported an error and the time elapsed between calibrating and testing. We assign a value of zero for the studies that test immediately after calibration. Our results show a relationship (SBP: $p=1.5e-5$ and DBP: $p=3.9e-7$) between the accuracy of personalization studies and Δt . However, the SBP and DBP estimation errors increase only by 0.763mmHg and 0.584mmHg for every 10-fold increase in seconds. The longest gap between calibration and test was two months [177].

Explained Deviations. We plot the SBP ED versus DBP ED for both calibration methods (Fig. 12). We found that studies that report high ED for SBP tend also to report high ED for DBP ($p=4.0e-18$). We delineate the ED_{min} in green. We observe that 22 subject split studies report EDs greater than the computed ED_{min} and none of the personalization studies would meet the standard. We further narrow the criteria to studies exceeding ED_{min} with 95% confidence. We report 9 studies that exceed ED_{min} with 95% confidence (Table. V-A). We observe that all studies are calibrated with subject split, the sensor data are mainly ECG+PPG and PPG (with one study that used ultrasound), and more than half of the studies use deep learning. However, these results are based on the data reported in the manuscripts, and the devices and algorithms will need further investigation and validation.

Commercially available devices. We report several commercially available devices (Table. V-A), namely BioBeat [180]–[182], OptiBP [183]–[188], Aktiia [189]–[197], HealthStats BPro or Medi-Watch [198]–[204], CheckMe [205]–[207], SomnoTouch [208], [209], Cloud DX Vitaliti [210], LiveMetric [211], and Valencell [212]. Valencell (FDA-approval pending) and LiveMetric (FDA-approved) show promising results. Valencell uses biometrics (age, weight, height, and gender) and PPG from finger and ear (as an earbud). LiveMetric uses biometrics (age, weight, and height) and waveform features from pressure sensors. It is important to note that the authors used mean band error instead of mean differences per ANSI/AAMI/ISO 81060-2:2019 Section 6.2.4 [60], which allows assigning errors as zero for estimates that are within the A-line reference range. This allowed the marketed error to drop from $0.2\pm10.1\text{mmHg}$ and $0.9\pm7.5\text{mmHg}$ to $0.0\pm6.9\text{mmHg}$ and $1.2\pm5.7\text{mmHg}$ for SBP and DBP, respectively. Performing this correction yields an ED of 3.51 and 2.49 for SBP and DBP.

B. Limitations

While only choosing to use ED is informative to investigate the heterogeneity in literature, its utility can be further improved by using traditional error metrics like STD of BP error and STD of BP or STD of BP change. We provide several cases below.

Case 1: low error and low ED. This indicates the dataset is too easy and, therefore, not helpful in evaluating algorithm/system performance. These studies tend to have small STDs for BP.

Case 2: low error and high ED. This indicates the algorithm is good. These studies tend to have representative or large STDs for BP. However, it may be possible to have small errors and small STDs for BP but high ED. Considering the BP and error distributions in analysis becomes important in this case.

Case 3: high error and low ED. This indicates the algorithm is not good.

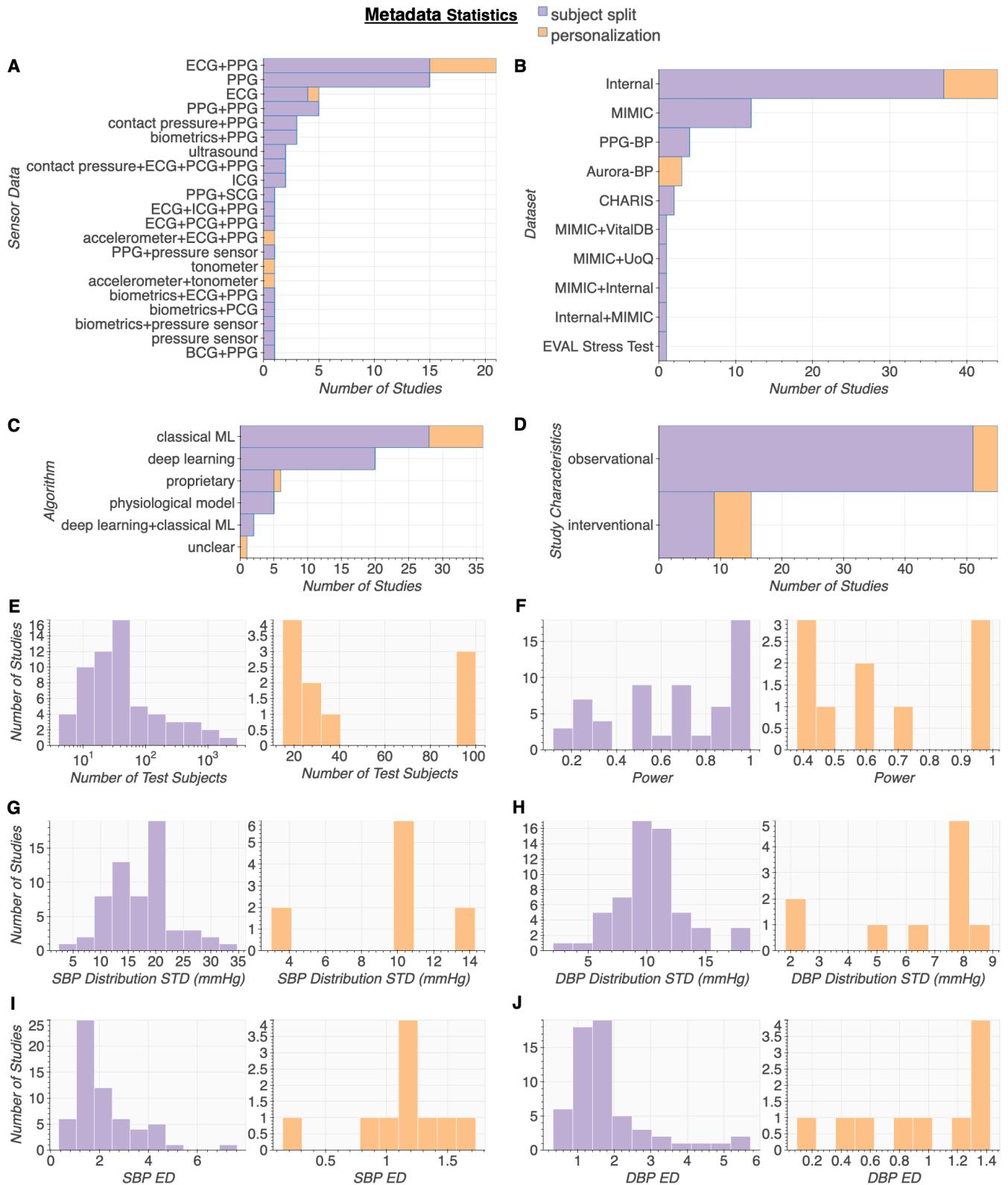
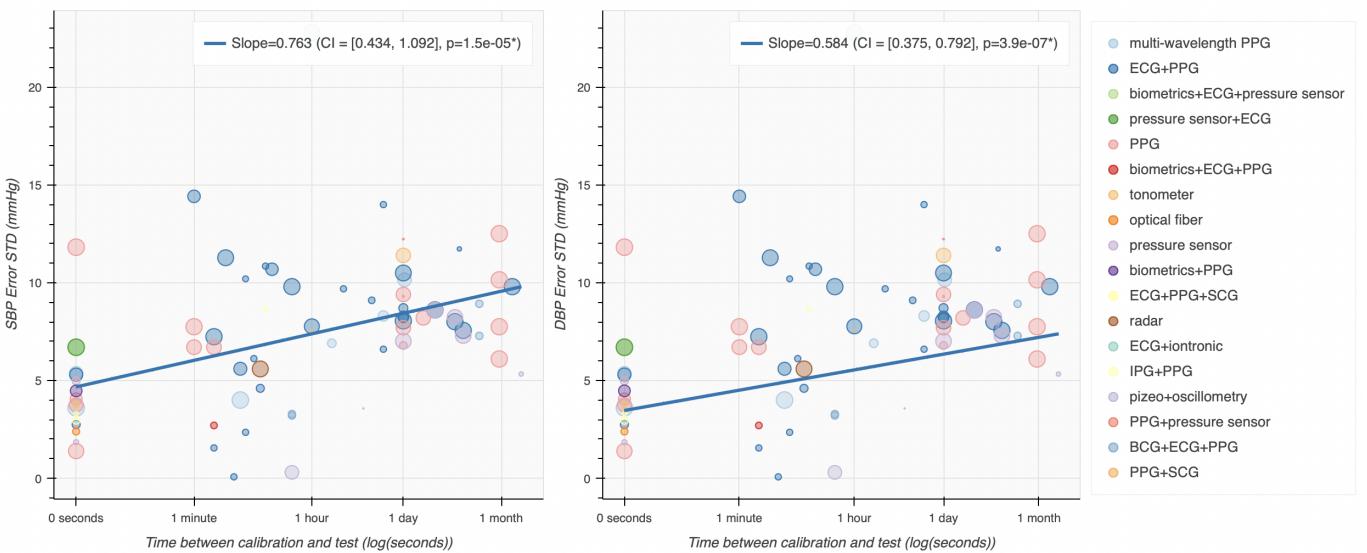
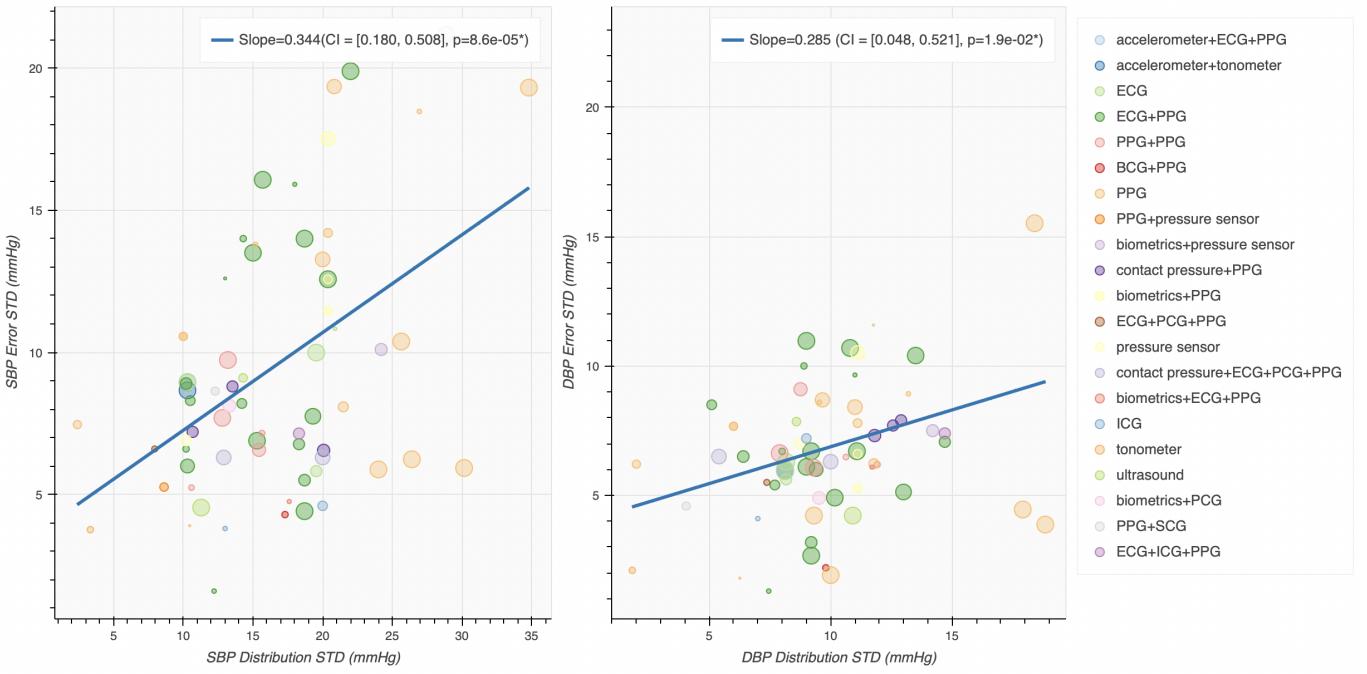


Fig. 9. Distribution of extracted parameters of studies included in the systematic review. Purple represents subject split studies, and orange represents the personalization studies. A. Sensors and systems: ECG and PPG+ECG are the most used sensors and systems. B. Datasets (internal or public): most datasets are internal and MIMIC is the most used publicly available dataset. C. Algorithm classes: classical ML and deep learning are the most used algorithm classes. D. Study characteristics (observational or interventional): most datasets are observational, and personalization studies are split between observational and interventional. E. Number of test subjects: most datasets contain less than 300 testing subjects, and the difference between subject split and personalization studies is insignificant. F. Study power: the difference between subject split and personalization studies is insignificant. G. and H. SBP and DBP STD: subject split studies tend to have larger STDs of BP compared to STDs of BP changes in personalization studies (SBP: $p=2.2e-2$ and DBP: $p=1.0e-2$). I. and J. SBP and DBP ED: subject split studies tend to have higher ED than personalization (SBP: $p=4.3e-6$ and DBP: $p=1.7e-4$). See <https://wearablebp.github.io/meta> for interactive figures.



Case 4: high error and high ED. This indicates that the dataset is extremely challenging and perhaps contains much more inter-subject variations than a representative sample.

VI. RESEARCH RECOMMENDATIONS

We strongly believe that the current wave of research activity and results in wearable blood pressure monitoring is going to have a significant impact and will likely result in providing us the ability to continuously monitor blood pressure, for the first time ever. This will

be transformational – leading to better clinical outcomes (especially for cardio-vascular diseases) and to better scientific understanding (especially on the mechanisms underlying how cardio-vascular health is related to other co-morbidities). While the pace of research in this area is indeed encouraging, we think that the lack of clearly defined experimental protocols, carefully-designed studies, and adequate standardization in the research setting is limiting the pace of innovation in the field. We provide several recommendations for the wearable BP research community that if adopted, can allow the community to better quantitatively compare across systems/algorithms enabling

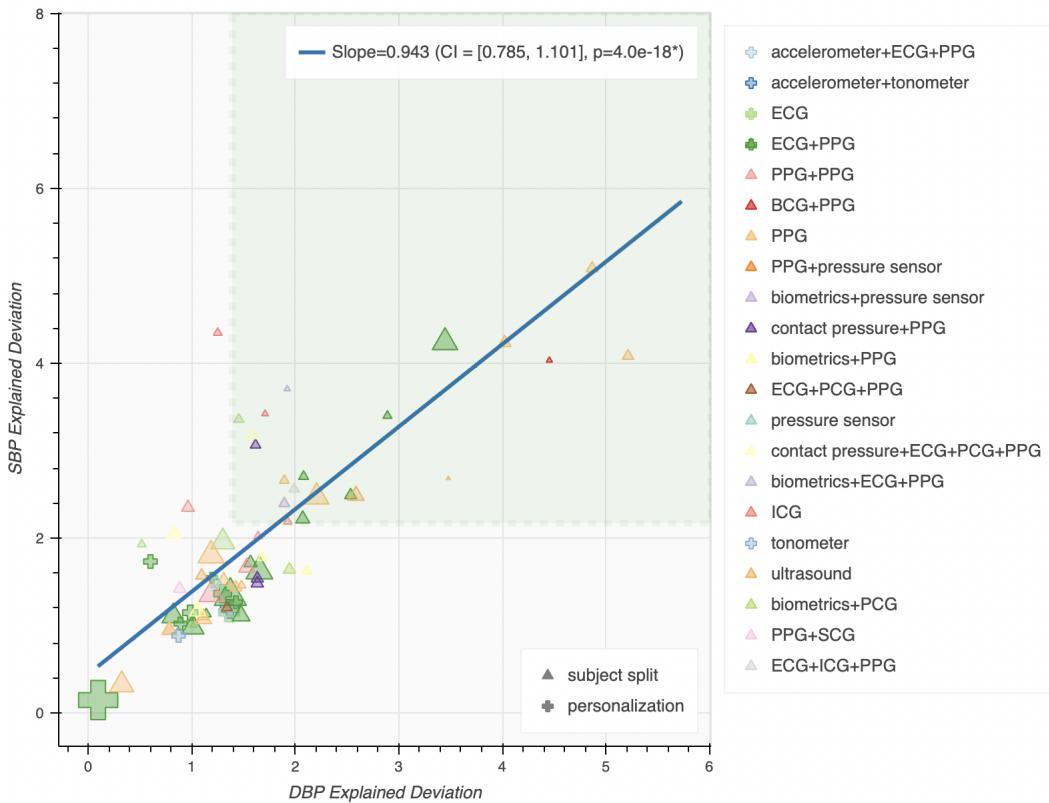


Fig. 12. Studies that report high Explained Deviation for SBP tend to also report high Explained Deviation for DBP. We show a scatter plot of DBP Explained Deviation versus SBP Explained Deviation. Here, the colors represent different sensor device configurations, the shape represents the calibration technique, and the size is inversely proportional to the sum of the SBP and DBP confidence intervals. We report the slope, its 95% confidence interval, and its corresponding p-value (* indicates significant: $p < 0.05$) in the legend inside the subplot. The green region bounds the minimum required Explained Deviations estimated from ANSI/AAMI/ISO 81060-2:2019 (See III). See <https://wearablebp.github.io/ed> for an interactive figure.

Title	Sensor Data	Algorithm	Dataset	SBP ED	DBP ED	SBP Distribution STD (mmHg)	DBP Distribution STD (mmHg)
Jeong (2021) [151]	ECG +PPG	deep learning	MIMIC	7.63	5.73	12.21	7.45
Ma (2022) [152]	PPG	deep learning	MIMIC	5.08	4.87	30.15	18.84
Huang (2022) [114]	ECG +PPG	deep learning	MIMIC	4.24	3.45	18.70	9.20
Ma (2022) [152]	PPG	deep learning	Internal	4.23	4.02	26.40	17.91
Dagamseh (2021) [143]	PPG	classical ML	Internal	4.08	5.21	24.00	10.00
Wu (2019) [178]	BCG +PPG	classical ML	Internal	4.03	4.45	17.30	9.80
Huang (2022) [114]	ECG +PPG	deep learning	MIMIC +UoQ	3.40	2.89	18.70	9.20
Hasanzadeh (2019) [146]	PPG	classical ML	MIMIC	2.47	2.21	25.64	9.31
Abolhasani (2020) [179]	ultrasound	physiological model	Internal	2.49	2.59	11.28	10.91

TABLE I

Studies that exceed ED_{min} with 95% confidence level. HERE, ALL THE STUDIES ARE CALIBRATED USING SUBJECT SPLIT. WE REPORT THE PERTINENT EXTRACTED PARAMETERS, NAMELY SENSOR DATA, ALGORITHM, DATASET, AND BP DISTRIBUTIONS, AND COMPUTE THE EXPLAINED DEVIATIONS (ED). NOTE THAT WHILE THESE STUDIES EXCEED THE ESTIMATED ED_{min} BASED ON THE STANDARDS, THEY MAY OR MAY NOT MEET THE MINIMUM BP DISTRIBUTION OR ERROR REQUIREMENTS. FURTHERMORE, THESE RESULTS ARE BASED ON THE DATA REPORTED IN THE MANUSCRIPTS, AND THE DEVICES AND/OR ALGORITHMS WILL NEED FURTHER INVESTIGATION AND VALIDATION.

future research to be focused on the most promising directions. This will in turn significantly accelerate the pace of innovation in this space resulting in better, practical, operational wearable BP monitoring systems.

- **Design studies with adequate BP variability and appropriate changes in BP distribution.** When assembling a dataset for

evaluation or for research development, it is essential that the dataset participants are a representative sample of the real world [8]. The majority of datasets reported in this systematic review were much more homogeneous (e.g., healthy young adults of college age) than the population as a whole. This results in two issues. Firstly, the standard deviation of BP or changes in BP

	Sensor(s)	Calibration Technique	SBP ED	DBP ED
BioBeat [180]–[182] (FDA-approved, CE certified)	Multi-wavelength PPG	Personalization	Inconclusive	Inconclusive
OptiBP [183]–[188]	PPG	Personalization	Inconclusive	Inconclusive
Akiita [189]–[197] (CE-marked Class IIA)	PPG	Personalization	Inconclusive	Inconclusive
HealthStats BPro [198]–[204] (previously known as MediWatch)	tonometer	Personalization	1.12 [198]	0.87 [198]
CheckMe [205]–[207]	ECG+PPG	Personalization	Inconclusive	Inconclusive
SomnoTouch [208], [209]	ECG+PPG	Personalization	Inconclusive	Inconclusive
CloudDX [210]	ECG+PPG	Personalization	Inconclusive	Inconclusive
LiveMetric (FDA-approved) [211]	Biometrics+ Pressure sensor	Subject Split	2.4 [211]	1.89 [211]
Valencell [212] (FDA-approval pending)	biometrics+ PPG	Subject Split	2.07 [212]	1.5 [212]

TABLE II

Commercially available devices. WE REPORT SENSORS, CALIBRATION TECHNIQUES, SBP ED, AND DBP ED FOR COMMERCIALLY AVAILABLE DEVICES. NOTABLY, LIVEMETRIC AND BIOBEAT ARE FDA-APPROVED, AND VALENCCELL IS PENDING FDA APPROVAL.

in the dataset itself is small, oftentimes below the ANSI/IEEE standard requirements for error – meaning even a blind estimator of population mean will look like it meets the standard. This has the effect of creating a mirage where many algorithms and systems meet standards while only a few would actually work in the real world – and diverts critical research resources into following many dead-ends. Second, when these methodologies and systems are finally implemented in real-world systems they do not generalize well and their performance predictably falls much below expectations – resulting in ‘buyers remorse’ which ultimately hurts the field. Moving forward, as a field, it is important that we collect data from a diverse cohort that mimics the real-world distribution.

- **Design studies with sufficient power.** Many datasets in wearable BP monitoring have less than 50 subjects and are not sufficiently powered to make statistically significant predictions. It is important that research studies rely on well-powered sample size. We understand that there are resource challenges with scaling sample sizes but we think there is an opportunity here for the community to collect a large-scale and public dataset that could be used to evaluate and benchmark algorithms.
- **Account for BP distribution when comparing across studies.** Even when datasets are well-powered and sufficiently diverse, direct quantitative comparisons across datasets are not appropriate. It is clear that the distribution of BP in a dataset is a significant confounder in reported BP estimation errors and must be accounted for appropriately before any quantitative comparisons are made. Accounting for BP distribution using explained deviation allows for a more reasonable (albeit still imperfect) quantitative comparison between studies.
- **Create public datasets.** The vast majority of existing studies, especially those done by commercial entities, are based on proprietary internal datasets, that are not public. In many cases, information about the BP distribution in the datasets is also unavailable, making it impossible to evaluate whether the study outcomes are significant or not. As a condition for publication, we should aim for transparency and insist that data is made public. Even if as a community, we do not want to go that far, we should insist that sufficient statistics regarding BP distribution in the dataset are available to at least make cursory evaluations of the potential impact of the study.
- **Create open-source software and common benchmarks.** As a research community, we should strive to release open-source common standardized algorithmic benchmarks that allow quantitative comparisons and will also serve to accelerate research outcomes. With this in mind, we have provided a first-of-its-

kind benchmark for wearable BP algorithms. The benchmark is open, and other researchers can add new algorithms and datasets to the benchmark allowing transparent research progress in the field.

- **Longitudinal studies for wearables in cardio-vascular health.**

We believe that sometime soon in the future, these devices will be robust enough to plan for a longitudinal study, that covers a sufficient number and diversity of enrolled subjects and seeks to measure the impact of wearable sensors for cardiovascular health – in particular, looking at how the availability of continuous measures changes subject behavior and potentially results in differences in outcomes. The lack of convincing clinical evidence for wearable monitoring systems is in part due to the lack of such a study – and this results in the current status quo where existing wearable devices are expensive, not covered by insurance and therefore practically only available for the well-resourced populations, further exacerbating health inequities in our society. A compelling study can be a powerful tool in ensuring that wearable devices are adopted within clinical practice as a standard of care, allowing these devices to be covered by insurance, thereby ensuring that all patients (somewhat independent of wealth) who might benefit from the technology have access to it.

APPENDIX I SYSTEMATIC REVIEW METHODOLOGY

A. Eligibility Criteria

Combining the three aforementioned study parameters (Calibration Technique, Explained Deviation, and Power), we developed Eligibility Criteria to determine whether studies should be included or excluded from the analysis. We include studies that:

- 1) Report a mean absolute error (MAE) or mean difference \pm standard deviation of the error (ME) for SBP and DBP estimation. We compute ME to MAE and vice versa (assuming zero bias) through a folded normal distribution. This is consistent with AAMI/ANSI/ISO 81060-2:2019 [60].
- 2) Written in the English language.
- 3) Claim to perform subject split or personalization.
- 4) Do not dynamically update based on individual subjects and their timestamps.
- 5) Perform experiments on human subjects, as we aim to deploy these devices on humans.
- 6) Report standard deviation for BP distribution in subject split studies and change in BP distribution for personalization studies. We also reported the time between calibration and test

		MIMIC [174]	PPG-BP [175]	VitalDB [176]
	Filter based on [147]	Filter based on [146]	Filter based on [175]	Filter based on [213]
Algorithm based on [143]	1.0:0.99	1.4:1.47	1.09:0.97	1.06:1.04
Algorithm based on [146]	1.05:1.02	1.45:1.4	1.13:1.0	1.05:1.08
Algorithm based on [151]	1.0:1.0	Not compatible	Not compatible	1.06:1.03
Algorithm based on [114]	1.0:1.0	Not compatible	Not compatible	1.05:1.08

TABLE III

RESULTS ($ED_{SBP} : ED_{DBP}$) FOR FOUR BENCHMARKED ALGORITHMS THAT REPORT HIGH ED, NAMELY ALGORITHMS PRESENTED IN, [146], [114], AND [151]. THESE ALGORITHMS WERE TRAINED AND TESTED ON MIMIC, PPG-BP, AND VITALDB FOR PPG (AND ECG) SENSOR DATA. WE IMPLEMENTED THE ALGORITHMS TO THE BEST OF OUR KNOWLEDGE AND WITH THE AVAILABLE CODE (IF IT EXISTS). AN UP-TO-DATE TABLE OF BENCHMARKS CAN BE FOUND AT [HTTPS://WEARABLEBP.GITHUB.IO/BENCHMARKS](https://wearablebp.github.io/benchmarks).

in personalization studies that do not have a change in BP distribution to attain Fig. 11.

- 7) Have test subjects greater than 1. The blood pressure distribution could not be computed if the number of subjects was one or less.

B. Information Sources and Search Strategy

The keywords “wearable”, “blood pressure”, “monitoring”, “estimation”, “systolic”, “diastolic”, and “cuffless” were entered into the search tool “Publish or Perish” [214] on 5/10/2022 which retrieves and analyzes academic citations from external data sources. We searched Crossref, Google Scholar, PubMed, Scopus, and Semantic Scholar. The maximum number of results was specified as 1000, and the articles were selected in the 1979-2022 time frame. Additional studies from Google Scholar alerts of “blood pressure device” and article references were included.

C. Selection Process

Whether a study met the inclusion criteria for review was determined by one individual (MC) who worked independently. To determine whether studies claim to perform subject split or personalization, MC manually looked for a combination of keywords and phrases. Examples include:

- Subject split: “used a separate x number of subjects for testing”, “did not mix training and testing subjects”, “separate training and testing subjects”
- Personalization: “used initial cuff value to calibrate model”, “calibrated model with cuff value”
- Record split without personalization: “shuffled training and testing segments”, “randomly sampled segments and assigned them to training and testing sets”, “overlapping segments”, “overlapping windows”

Studies with unclear calibration techniques were excluded. Personalization studies that reported testing the device immediately after calibration were assigned a time between calibration and test of 0 seconds.

D. Data Collection Process and Data Items

One individual (MC) extracted predefined data parameters. For papers that had multiple protocols and reported multiple results, MC made multiple entries. The predefined data fields include Key Devices and Measurements, Calibration Technique, Algorithm, Dataset, Number of Test Subjects, Training Subject Characteristics, Testing Subject Characteristics, Study Characteristics (observational or interventional study), BP Distribution, Evaluation Metric, and Reported Result. MC reported the time between calibrating and testing the model for personalization articles. Moreover, MC used (if available) figures with an axis with Reference SBP or DBP values (correlation plots or reference versus error plots) to estimate the SBP and DBP population

standard deviation using image processing techniques when the BP distribution was not provided in the code. To determine whether the sample size is sufficient to detect the reported result, MC computed study power $p = 1 - \beta = \phi(ES\sqrt{\frac{n}{2}} - z_{1-\alpha/2})$ based on ANSI/AAMI/ISO 81060-2:2019 [60] by defining the probability of type 1 error as $\alpha = 0.05$, probability of type 2 error as $\beta = 0.02$, effect size (ES) as the difference in means divided by standard deviation $ES = \frac{5}{8}$, and ϕ is the cumulative distribution function of a normal distribution. A power of $p = 0.98$ corresponds to a sample size of approximately 84. Finally, MC reported implementation and dataset availability for studies included in the analysis. The raw data is available upon request.

E. Study Risk of Bias Assessment and Reporting Bias Assessment

Since one individual (MC) decided whether studies were included and extracted all study parameters, there is an inherent risk of bias. MC reviewed all the included studies again to determine whether there were any errors.

Studies that reported the standard deviation of the (change in) BP distribution in the study sample were included. However, this excluded potentially promising studies that reported BP distribution ranges for subject split studies or change in BP ranges for personalization studies. Similarly, promising studies that did not report MAE or ME were potentially excluded. These studies tended to report metrics such as root mean square error (RMSE).

APPENDIX II EXPLAINED DEVIATION HYPOTHESIS TESTING

Statistically, ED is similar to an F-test for the ratio of two variances with equal sample sizes. As a consequence, we can construct a null hypothesis $H_0 : ED \leq ED_{min}$ and the alternative hypothesis $H_1 : ED > ED_{min}$ where ED_{min} is the computed minimum ED that meets the requirements from the standard. We can compute confidence intervals by determining the bounds:

$$P(\sqrt{F_{1-\alpha/2}(n-1, n-1)}ED_{est} \leq ED_{true} \leq \sqrt{F_{\alpha/2}(n-1, n-1)}ED_{est}) = 1 - \alpha$$

where α is the level of significance, P is the probability, F is the F-distribution, and n in the sample size. To derive this expression, assume:

- 1) n_1 independent samples from a normally distributed population with variance σ_1^2
- 2) n_2 independent samples from a normally distributed population with variance σ_2^2

Given sample variance s_1^2 and s_2^2 , we can write $\frac{s_1^2}{\sigma_1^2} \sim \chi_{df_1}^2$ and $\frac{s_2^2}{\sigma_2^2} \sim \chi_{df_2}^2$. In this case, $df_1 = n - 1$ and $df_2 = n - 1$. The ratio of

these two distributions is an F-distribution and can be written in the form:

$$F \sim \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

The $1 - \alpha$ confidence interval can be written as

$$\begin{aligned} P(F_{\alpha/2}(n-1, n-1) \leq \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \leq F_{1-\alpha/2}(n-1, n-1)) &= 1 - \alpha \\ \text{Substituting } \frac{s_2}{\sigma_1} = ED_{est} \text{ and } \frac{\sigma_2}{\sigma_1} = ED_{true} \text{ gives} \end{aligned}$$

$$\begin{aligned} P(\sqrt{F_{1-\alpha/2}(n-1, n-1)}ED_{est} \leq ED_{true} \leq \sqrt{F_{\alpha/2}(n-1, n-1)}ED_{est}) &= 1 - \alpha \end{aligned}$$

APPENDIX III MINIMUM EXPLAINED DEVIATION BASED ON ANSI/AAMI/ISO 81060-2:2019

We computed the minimum Explained Deviation (ED_{min}) as 2.17 and 1.39 based on ANSI/AAMI/ISO 81060-2:2019 for SBP and DBP. We used the PPG-BP database from [175] to compute SBP and DBP standard deviations for sample size 85 using weighted sampling based on an iterative proportional fitting procedure. Code for this procedure can be found at <https://github.com/wearablebp/wearablebp-benchmarks>. Then, we iterated this procedure 10000 times to find the minimum dataset SBP and DBP standard deviation. Note that ED_{min} for SBP is larger than for DBP because the error requirements for SBP and DBP are the same, but the range for DBP on the high end above 120/80 mmHg (20% at 85 mmHg and 5% at 100 mmHg) is much smaller than that for SBP (5% at 160 mmHg and 20% at 140 mmHg).

APPENDIX IV BENCHMARK RESULTS

We benchmark algorithms from [146], [114], [143], and [151] on public datasets with sufficient power, namely MIMIC [174], PPG-BP [175], and VitalDB [176]. We report ED for SBP and DBP. See Table III.

REFERENCES

- [1] W. H. Organization *et al.*, "World health statistics: a snapshot of global health," Organización Mundial de la Salud, Tech. Rep., 2012.
- [2] C. for Disease Control, P. (CDC) *et al.*, "Hypertension cascade: hypertension prevalence, treatment and control estimates among us adults aged 18 years and older applying the criteria from the american college of cardiology and american heart association's 2017 hypertension guideline—nhanes 2013–2016," Atlanta, GA: US Department of Health and Human Services, 2019.
- [3] D. Shimbo, N. T. Artinian, J. N. Basile, L. R. Krakoff, K. L. Margolis, M. K. Rakotz, G. Wozniak, A. H. Association, and the American Medical Association, "Self-measured blood pressure monitoring at home: a joint policy statement from the american heart association and american medical association," *Circulation*, vol. 142, no. 4, pp. e42–e63, 2020.
- [4] J. George and T. MacDonald, "Home blood pressure monitoring," *European Cardiology Review*, vol. 10, no. 2, p. 95, 2015.
- [5] E. O'Brien, K. Kario, J. A. Staessen, A. de la Sierra, and T. Ohkubo, "Patterns of ambulatory blood pressure: clinical relevance and application," *The Journal of Clinical Hypertension*, vol. 20, no. 7, pp. 1112–1115, 2018.
- [6] S. J. Mann, "The clinical spectrum of labile hypertension: a management dilemma," *The Journal of Clinical Hypertension*, vol. 11, no. 9, pp. 491–497, 2009.
- [7] L. Zhao, C. Liang, Y. Huang, G. Zhou, Y. Xiao, N. Ji, Y.-T. Zhang, and N. Zhao, "Emerging sensing and modeling technologies for wearable and cuffless blood pressure monitoring," *npj Digital Medicine*, vol. 6, no. 1, p. 93, 2023.
- [8] R. Mukkamala, G. S. Stergiou, and A. P. Avolio, "Cuffless blood pressure measurement," *Annual Review of Biomedical Engineering*, vol. 24, no. 1, pp. 203–230, 2022.
- [9] M. Elgendi, R. Fletcher, Y. Liang, N. Howard, N. H. Lovell, D. Abbott, K. Lim, and R. Ward, "The use of photoplethysmography for assessing hypertension," *NPJ digital medicine*, vol. 2, no. 1, p. 60, 2019.
- [10] R. Mukkamala, J.-O. Hahn, and A. Chandrasekhar, "Photoplethysmography in noninvasive blood pressure monitoring," in *Photoplethysmography*. Elsevier, 2022, pp. 359–400.
- [11] K. B. Kim and H. J. Baek, "Photoplethysmography in wearable devices: A comprehensive review of technological advances, current challenges, and future directions," *Electronics*, vol. 12, no. 13, p. 2923, 2023.
- [12] M. Hosanee, G. Chan, K. Welykholowa, R. Cooper, P. A. Kyriacou, D. Zheng, J. Allen, D. Abbott, C. Menon, N. H. Lovell *et al.*, "Cuffless single-site photoplethysmography for blood pressure monitoring," *Journal of clinical medicine*, vol. 9, no. 3, p. 723, 2020.
- [13] J.-R. Hu, G. Martin, S. Iyengar, L. C. Kovell, T. B. Plante, N. van Helmond, R. A. Dart, T. M. Brady, R.-A. N. Turkson-Ocran, and S. P. Jurasichek, "Validating cuffless continuous blood pressure monitoring devices," *Cardiovascular Digital Health Journal*, 2023.
- [14] R. Mukkamala, M. Yavarimanesh, K. Natarajan, J.-O. Hahn, K. G. Kyriakoulis, A. P. Avolio, and G. S. Stergiou, "Evaluation of the accuracy of cuffless blood pressure measurement devices: challenges and proposals," *Hypertension*, vol. 78, no. 5, pp. 1161–1167, 2021.
- [15] N. Pilz, A. Patzak, and T. L. Bothe, "Continuous cuffless and non-invasive measurement of arterial blood pressure—concepts and future perspectives," *Blood Pressure*, vol. 31, no. 1, pp. 254–269, 2022.
- [16] C. K. Bradley, D. Shimbo, D. A. Colburn, D. N. Pugliese, R. Padwal, S. K. Sia, and D. E. Anstey, "Cuffless blood pressure devices," *American Journal of Hypertension*, vol. 35, no. 5, pp. 380–387, 2022.
- [17] T. Panula, J.-P. Sirkiä, D. Wong, and M. Kaisti, "Advances in non-invasive blood pressure measurement techniques," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 424–438, 2022.
- [18] C. El-Hajj and P. A. Kyriacou, "A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure," *Biomedical Signal Processing and Control*, vol. 58, p. 101870, 2020.
- [19] N. D. Agham and U. M. Chaskar, "Learning and non-learning algorithms for cuffless blood pressure measurement: a review," *Medical & Biological Engineering & Computing*, vol. 59, no. 6, pp. 1201–1222, 2021.
- [20] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.
- [21] J. Fine, K. L. Branan, A. J. Rodriguez, T. Boonya-Ananta, J. C. Ramella-Roman, M. J. McShane, and G. L. Coté, "Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring," *Biosensors*, vol. 11, no. 4, p. 126, 2021.
- [22] W. Zijlstra, A. Buursma, and W. Meeuwesen-Van der Roest, "Absorption spectra of human fetal and adult oxyhemoglobin, de-oxyhemoglobin, carboxyhemoglobin, and methemoglobin," *Clinical chemistry*, vol. 37, no. 9, pp. 1633–1638, 1991.
- [23] J. Lee, K. Matsumura, K.-i. Yamakoshi, P. Rolfe, S. Tanaka, and T. Yamakoshi, "Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2013, pp. 1724–1727.
- [24] W. Cui, L. E. Ostrander, and B. Y. Lee, "In vivo reflectance of blood and tissue as a function of light wavelength," *IEEE transactions on biomedical engineering*, vol. 37, no. 6, pp. 632–639, 1990.
- [25] S. Chatterjee, K. Budidha, and P. A. Kyriacou, "Investigating the origin of photoplethysmography using a multiwavelength monte carlo model," *Physiological Measurement*, vol. 41, no. 8, p. 084001, 2020.
- [26] Y. Mendelson and B. D. Ochs, "Noninvasive pulse oximetry utilizing skin reflectance photoplethysmography," *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 10, pp. 798–805, 1988.
- [27] T. Boonya-Ananta, A. J. Rodriguez, A. Ajmal, V. N. Du Le, A. K. Hansen, J. D. Hutcheson, and J. C. Ramella-Roman, "Synthetic photoplethysmography (ppg) of the radial artery through parallelized monte carlo and its correlation to body mass index (bmi)," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [28] G. Zonios, J. Bykowski, and N. Kollias, "Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed in vivo

- using diffuse reflectance spectroscopy,” *Journal of Investigative Dermatology*, vol. 117, no. 6, pp. 1452–1457, 2001.
- [29] K. Sel, D. Osman, and R. Jafari, “Non-invasive cardiac and respiratory activity assessment from various human body locations using bioimpedance,” *IEEE open journal of engineering in medicine and biology*, vol. 2, pp. 210–217, 2021.
- [30] L. E. Baker, “Principles of the impedance technique,” *IEEE engineering in medicine and biology magazine*, vol. 8, no. 1, pp. 11–15, 1989.
- [31] K. Sel, D. Osman, N. Huerta, A. Edgar, R. I. Pettigrew, and R. Jafari, “Continuous cuffless blood pressure monitoring with a wearable ring bioimpedance device,” *npj Digital Medicine*, vol. 6, no. 1, p. 59, 2023.
- [32] B. Ibrahim and R. Jafari, “Cuffless blood pressure monitoring from a wristband with calibration-free algorithms for sensing location based on bio-impedance sensor array and autoencoder,” *Scientific reports*, vol. 12, no. 1, pp. 1–14, 2022.
- [33] D. Kireev, K. Sel, B. Ibrahim, N. Kumar, A. Akbari, R. Jafari, and D. Akinwande, “Continuous cuffless monitoring of arterial blood pressure via graphene bioimpedance tattoos,” *Nature Nanotechnology*, vol. 17, no. 8, pp. 864–870, 2022.
- [34] W. Lee and Y. Roh, “Ultrasonic transducers for medical diagnostic imaging,” *Biomedical engineering letters*, vol. 7, no. 2, pp. 91–97, 2017.
- [35] G. York and Y. Kim, “Ultrasound processing and computing: Review and future directions,” *Annual review of biomedical engineering*, vol. 1, no. 1, pp. 559–588, 1999.
- [36] C. Basoglu, R. Managuli, G. York, and Y. Kim, “Computing requirements of modern medical diagnostic ultrasound machines,” *Parallel Computing*, vol. 24, no. 9–10, pp. 1407–1431, 1998.
- [37] A. Ramalli, E. Boni, E. Roux, H. Liebgott, and P. Tortoli, “Design, implementation, and medical applications of 2-d ultrasound sparse arrays,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 10, pp. 2739–2755, 2022.
- [38] Y. Baribeau, A. Sharkey, O. Chaudhary, S. Krumm, H. Fatima, F. Mahmood, and R. Matyal, “Handheld point-of-care ultrasound probes: the new generation of pocus,” *Journal of cardiothoracic and vascular anesthesia*, vol. 34, no. 11, pp. 3139–3145, 2020.
- [39] J. R. Sempionatto, M. Lin, L. Yin, E. De la Paz, K. Pei, T. Sonsa-Ard, A. N. de Loyola Silva, A. A. Khorshed, F. Zhang, N. Tostado *et al.*, “An epidermal patch for the simultaneous monitoring of haemodynamic and metabolic biomarkers,” *Nature Biomedical Engineering*, vol. 5, no. 7, pp. 737–748, 2021.
- [40] H. Hu, H. Huang, M. Li, X. Gao, L. Yin, R. Qi, R. S. Wu, X. Chen, Y. Ma, K. Shi *et al.*, “A wearable cardiac ultrasound imager,” *Nature*, vol. 613, no. 7945, pp. 667–675, 2023.
- [41] C. Wang, X. Li, H. Hu, L. Zhang, Z. Huang, M. Lin, Z. Zhang, Z. Yin, B. Huang, H. Gong *et al.*, “Monitoring of the central blood pressure waveform via a conformal ultrasonic device,” *Nature biomedical engineering*, vol. 2, no. 9, pp. 687–695, 2018.
- [42] K. Meng, X. Xiao, W. Wei, G. Chen, A. Nashalian, S. Shen, X. Xiao, and J. Chen, “Wearable pressure sensors for pulse wave monitoring,” *Advanced Materials*, vol. 34, no. 21, p. 2109357, 2022.
- [43] C.-S. Kim, A. M. Carek, O. T. Inan, R. Mukkamala, and J.-O. Hahn, “Ballistocardiogram-based approach to cuffless blood pressure monitoring: Proof of concept and potential challenges,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2384–2391, 2018.
- [44] C.-S. Kim, A. M. Carek, R. Mukkamala, O. T. Inan, and J.-O. Hahn, “Ballistocardiogram as proximal timing reference for pulse transit time measurement: Potential for cuffless blood pressure monitoring,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 11, pp. 2657–2664, 2015.
- [45] J. Qin, L.-J. Yin, Y.-N. Hao, S.-L. Zhong, D.-L. Zhang, K. Bi, Y.-X. Zhang, Y. Zhao, and Z.-M. Dang, “Flexible and stretchable capacitive sensors with different microstructures,” *Advanced Materials*, vol. 33, no. 34, p. 2008267, 2021.
- [46] M. Amjadi, K.-U. Kyung, I. Park, and M. Sitti, “Stretchable, skin-mountable, and wearable strain sensors and their potential applications: a review,” *Advanced Functional Materials*, vol. 26, no. 11, pp. 1678–1698, 2016.
- [47] T. Kleckers, “Force sensors for strain gauge and piezoelectric crystal-based mechatronic systems-a comparison,” in *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. IEEE, 2012, pp. 2306–2308.
- [48] J. Fortin, D. E. Rogge, C. Fellner, D. Flotzinger, J. Grond, K. Lerche, and B. Saugel, “A novel art of continuous noninvasive blood pressure measurement,” *Nature communications*, vol. 12, no. 1, pp. 1–14, 2021.
- [49] J. Penaz, “Photoelectric measurement of blood pressure, volume and flow in finger,” *Dig 10th ICMBE*, vol. 104, 1973.
- [50] B. P. IMHOLZ, G. A. V. MONTFRANS, J. J. SETTELS, G. M. V. D. HOEVEN, J. M. KAREMAKER, and W. WIELING, “Continuous non-invasive blood pressure monitoring: reliability of finapres device during the valsalva manoeuvre,” *Cardiovascular research*, vol. 22, no. 6, pp. 390–397, 1988.
- [51] K. Wesseling, “Finger arterial pressure measurement with finapres.” *Zeitschrift für Kardiologie*, vol. 85, pp. 38–44, 1996.
- [52] R. Maestri, G. Pinna, E. Robbi, S. Capomolla, and M. La Rovere, “Noninvasive measurement of blood pressure variability: accuracy of the finometer monitor and comparison with the finapres device,” *Physiological measurement*, vol. 26, no. 6, p. 1125, 2005.
- [53] B. Silke and D. McAuley, “Accuracy and precision of blood pressure determination with the finapres: an overview using re-sampling statistics,” *Journal of human hypertension*, vol. 12, no. 6, pp. 403–409, 1998.
- [54] A. Chandrasekhar, C.-S. Kim, M. Naji, K. Natarajan, J.-O. Hahn, and R. Mukkamala, “Smartphone-based blood pressure monitoring via the oscillometric finger-pressing method,” *Science translational medicine*, vol. 10, no. 431, p. eaap8674, 2018.
- [55] M. Freithaler, A. Chandrasekhar, V. Dhamotharan, C. Landry, S. G. Shroff, and R. Mukkamala, “Smartphone-based blood pressure monitoring via the oscillometric finger pressing method: Analysis of oscillation width variations can improve diastolic pressure computation,” *IEEE Transactions on Biomedical Engineering*, 2023.
- [56] Y. Xuan, C. Barry, J. De Souza, J. H. Wen, N. Antipa, A. A. Moore, and E. J. Wang, “Ultra-low-cost mechanical smartphone attachment for no-calibration blood pressure measurement,” *Scientific Reports*, vol. 13, no. 1, p. 8105, 2023.
- [57] Y. Cao, H. Chen, F. Li, and Y. Wang, “Crisp-bp: Continuous wrist ppg-based blood pressure measurement,” in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 378–391.
- [58] P. L. Mitchell, R. W. Parlin, and H. Blackburn, “Effect of vertical displacement of the arm on indirect blood-pressure measurement,” *New England Journal of Medicine*, vol. 271, no. 2, pp. 72–74, 1964.
- [59] D. Silverberg, E. Shemesh, and A. Iaina, “The unsupported arm: a cause of falsely raised blood pressure readings.” *British Medical Journal*, vol. 2, no. 6098, p. 1331, 1977.
- [60] “Non-invasive sphygmomanometers— Part 2: Clinical investigation of intermittent automated measurement type,” American National Standard, Standard, Mar. 2019.
- [61] P. A. Shaltis, A. T. Reisner, and H. H. Asada, “Cuffless blood pressure monitoring using hydrostatic pressure changes,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 6, pp. 1775–1777, 2008.
- [62] M. Hickey, J. P. Phillips, and P. A. Kyriacou, “Investigation of peripheral photoplethysmographic morphology changes induced during a hand-elevation study,” *Journal of clinical monitoring and computing*, vol. 30, pp. 727–736, 2016.
- [63] X. Jiang, S. Wei, D. Zheng, F. Liu, S. Zhang, Z. Zhang, and C. Liu, “Change of bilateral difference in radial artery pulse morphology with one-side arm movement,” *Artery Research*, vol. 19, pp. 1–8, 2017.
- [64] G. Pucci, F. Battista, F. Anastasio, L. Sanesi, B. Gavish, M. Butlin, A. Avolio, and G. Schillaci, “Effects of gravity-induced upper-limb blood pressure changes on wave transmission and arterial radial waveform,” *Journal of hypertension*, vol. 34, no. 6, pp. 1091–1098, 2016.
- [65] C. C. Poon, Y.-T. Zhang, and Y. Liu, “Modeling of pulse transit time under the effects of hydrostatic pressure for cuffless blood pressure measurements,” in *2006 3rd IEEE/EMBS International Summer School on Medical Devices and Biosensors*. IEEE, 2006, pp. 65–68.
- [66] S. McCombieD, “Adaptive hydrostatic blood pressure calibration: development of a wearable, autonomous pulse wave velocity blood pressure monitor,” *11th Proceedings of the 29th Annual International Conference of the IEEE EMBS. Lyon: IEEE*, 2007.
- [67] P. Shaltis, A. Reisner, and H. Asada, “A hydrostatic pressure approach to cuffless blood pressure monitoring,” in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1. IEEE, 2004, pp. 2173–2176.
- [68] P. A. Shaltis, A. Reisner, and H. H. Asada, “Wearable, cuff-less ppg-based blood pressure monitor with novel height sensor,” in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 908–911.
- [69] C. C. Poon and Y.-T. Zhang, “Using the changes in hydrostatic pressure and pulse transit time to measure arterial blood pressure,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 2336–2337.
- [70] H. Sato, H. Koshimizu, S. Yamashita, and T. Ogura, “Blood pressure monitor with a position sensor for wrist placement to eliminate hydrostatic pressure effect on blood pressure measurement,” in *2013 35th*

- Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 1835–1838.
- [71] B. Gavish and L. Gavish, “Blood pressure variation in response to changing arm cuff height cannot be explained solely by the hydrostatic effect,” *Journal of hypertension*, vol. 29, no. 11, pp. 2099–2104, 2011.
- [72] R. Mukkamala, J.-O. Hahn, O. T. Inan, L. K. Mestha, C.-S. Kim, H. Töreyin, and S. Kyal, “Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 1879–1901, 2015.
- [73] D. Franklin, A. Tzavelis, J. Y. Lee, H. U. Chung, J. Trueb, H. Arafa, S. S. Kwak, I. Huang, Y. Liu, M. Rathod *et al.*, “Synchronized wearables for the detection of haemodynamic states via electrocardiography and multispectral photoplethysmography,” *Nature biomedical engineering*, pp. 1–13, 2023.
- [74] J. Liu, B. P.-Y. Yan, W.-X. Dai, X.-R. Ding, Y.-T. Zhang, and N. Zhao, “Multi-wavelength photoplethysmography method for skin arterial pulse extraction,” *Biomedical optics express*, vol. 7, no. 10, pp. 4313–4326, 2016.
- [75] J. Liu, S. Qiu, N. Luo, S.-K. Lau, H. Yu, T. Kwok, Y.-T. Zhang, and N. Zhao, “Pca-based multi-wavelength photoplethysmography algorithm for cuffless blood pressure measurement on elderly subjects,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 663–673, 2020.
- [76] D. Ray, T. Collins, S. Woolley, and P. Ponnappalli, “A review of wearable multi-wavelength photoplethysmography,” *IEEE Reviews in Biomedical Engineering*, 2021.
- [77] S. S. Franklin, W. Gustin IV, N. D. Wong, M. G. Larson, M. A. Weber, W. B. Kannel, and D. Levy, “Hemodynamic patterns of age-related changes in blood pressure: the framingham heart study,” *Circulation*, vol. 96, no. 1, pp. 308–315, 1997.
- [78] J. Liu, Y. Li, X.-R. Ding, W.-X. Dai, and Y.-T. Zhang, “Effects of cuff inflation and deflation on pulse transit time measured from ecg and multi-wavelength ppg,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 5973–5976.
- [79] J. Liu, B. P. Yan, Y.-T. Zhang, X.-R. Ding, P. Su, and N. Zhao, “Multi-wavelength photoplethysmography enabling continuous blood pressure measurement with compact wearable electronics,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 6, pp. 1514–1525, 2018.
- [80] J. Giltvedt, A. Sira, and P. Helme, “Pulsed multifrequency photoplethysmograph,” *Medical and Biological Engineering and Computing*, vol. 22, no. 3, pp. 212–215, 1984.
- [81] A. Chandrasekhar, M. Yavarimanesh, K. Natarajan, J.-O. Hahn, and R. Mukkamala, “Ppg sensor contact pressure should be taken into account for cuff-less blood pressure measurement,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3134–3140, 2020.
- [82] J. Fine, M. J. McShane, G. L. Coté, and C. G. Scully, “A computational modeling and simulation workflow to investigate the impact of patient-specific and device factors on hemodynamic measurements from non-invasive photoplethysmography,” *Biosensors*, vol. 12, no. 8, p. 598, 2022.
- [83] A. J. Rodriguez, M. T. Boonya-Ananta, M. Gonzalez, V. N. D. Le, J. Fine, C. Palacios, M. J. McShane, G. L. Coté, and J. C. Ramella-Roman, “Skin optical properties in the obese and their relation to body mass index: a review,” *Journal of Biomedical Optics*, vol. 27, no. 3, pp. 030902–030902, 2022.
- [84] T.-W. Wang, H.-W. Chu, W.-X. Chen, Y.-T. Shih, P.-C. Hsu, H.-M. Cheng, and S.-F. Lin, “Single-channel impedance plethysmography neck patch device for unobtrusive wearable cardiovascular monitoring,” *IEEE Access*, vol. 8, pp. 184909–184919, 2020.
- [85] V. P. Rachim, T. H. Huynh, and W.-Y. Chung, “Wrist photoplethysmography and bio-impedance sensor for cuff-less blood pressure monitoring,” in *2018 Ieee Sensors*. IEEE, 2018, pp. 1–4.
- [86] P. Nabeel, J. Jayaraj, K. Srinivasa, S. Mohanasankar, and M. Chenniappan, “Bi-modal arterial compliance probe for calibration-free cuffless blood pressure estimation,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2392–2404, 2018.
- [87] Y. Yoon, J. H. Cho, and G. Yoon, “Non-constrained blood pressure monitoring using ecg and ppg for personal healthcare,” *Journal of medical systems*, vol. 33, pp. 261–266, 2009.
- [88] R. Payne, C. Symeonides, D. Webb, and S. Maxwell, “Pulse transit time measured from the ecg: an unreliable marker of beat-to-beat blood pressure,” *Journal of Applied Physiology*, vol. 100, no. 1, pp. 136–141, 2006.
- [89] M. Simjanoska, M. Gjoreski, M. Gams, and A. Madevska Bogdanova, “Non-invasive blood pressure estimation from ecg using machine learning techniques,” *Sensors*, vol. 18, no. 4, p. 1160, 2018.
- [90] R. P. Smith, J. Argod, J.-L. Pépin, and P. A. Lévy, “Pulse transit time: an appraisal of potential clinical applications,” *Thorax*, vol. 54, no. 5, pp. 452–457, 1999.
- [91] E. Finnegan, S. Davidson, M. Harford, J. Jorge, P. Watkinson, D. Young, L. Tarassenko, and M. Villarroel, “Pulse arrival time as a surrogate of blood pressure,” *Scientific reports*, vol. 11, no. 1, pp. 1–21, 2021.
- [92] A. Taebi, B. E. Solar, A. J. Bomar, R. H. Sandler, and H. A. Mansy, “Recent advances in seismocardiography,” *Vibration*, vol. 2, no. 1, pp. 64–86, 2019.
- [93] A. M. Carek, J. Conant, A. Joshi, H. Kang, and O. T. Inan, “Seismowatch: wearable cuffless blood pressure monitoring using pulse transit time,” *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 1, no. 3, pp. 1–16, 2017.
- [94] E. J. Wang, J. Zhu, M. Jain, T.-J. Lee, E. Saba, L. Nachman, and S. N. Patel, “Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera,” in *Proceedings of the 2018 CHI conference on human factors in computing Systems*, 2018, pp. 1–9.
- [95] V. G. Ganti, A. M. Carek, B. N. Nevius, J. A. Heller, M. Etemadi, and O. T. Inan, “Wearable cuff-less blood pressure estimation at home via pulse transit time,” *IEEE journal of biomedical and health informatics*, vol. 25, no. 6, pp. 1926–1937, 2020.
- [96] A. E. Dastjerdi, M. Kachuee, and M. Shabany, “Non-invasive blood pressure estimation using phonocardiogram,” in *2017 IEEE international symposium on circuits and systems (ISCAS)*. IEEE, 2017, pp. 1–4.
- [97] T. Omari and F. Berekshi-Reguig, “A new approach for blood pressure estimation based on phonocardiogram,” *Biomedical engineering letters*, vol. 9, pp. 395–406, 2019.
- [98] H. U. Chung, A. Y. Rwei, A. Hourlier-Fargette, S. Xu, K. Lee, E. C. Dunne, Z. Xie, C. Liu, A. Carlini, D. H. Kim *et al.*, “Skin-interfaced biosensors for advanced wireless physiological monitoring in neonatal and pediatric intensive-care units,” *Nature medicine*, vol. 26, no. 3, pp. 418–429, 2020.
- [99] S. Chen, J. Qi, S. Fan, Z. Qiao, J. C. Yeo, and C. T. Lim, “Flexible wearable sensors for cardiovascular health monitoring,” *Advanced Healthcare Materials*, vol. 10, no. 17, p. 2100116, 2021.
- [100] Y. Luo, M. R. Abidian, J.-H. Ahn, D. Akinwande, A. M. Andrews, M. Antonietti, Z. Bao, M. Berggren, C. A. Berkey, C. J. Bettinger *et al.*, “Technology roadmap for flexible sensors,” *ACS nano*, vol. 17, no. 6, pp. 5211–5295, 2023.
- [101] C. Vlachopoulos, M. O’Rourke, and W. W. Nichols, *McDonald’s blood flow in arteries: theoretical, experimental and clinical principles*. CRC press, 2011.
- [102] S. Huthart, M. Elgendi, D. Zheng, G. Stansby, and J. Allen, “Advancing ppg signal quality and know-how through knowledge translation—from experts to student and researcher,” *Frontiers in Digital Health*, vol. 2, p. 619692, 2020.
- [103] M. Elgendi, “On the analysis of fingertip photoplethysmogram signals,” *Current cardiology reviews*, vol. 8, no. 1, pp. 14–25, 2012.
- [104] Y. Liang, M. Elgendi, Z. Chen, and R. Ward, “An optimal filter for short photoplethysmogram signals,” *Scientific data*, vol. 5, no. 1, pp. 1–12, 2018.
- [105] M. Elgendi, “Optimal signal quality index for photoplethysmogram signals,” *Bioengineering*, vol. 3, no. 4, p. 21, 2016.
- [106] Q. Li and G. D. Clifford, “Dynamic time warping and machine learning for signal quality assessment of pulsatile signals,” *Physiological measurement*, vol. 33, no. 9, p. 1491, 2012.
- [107] W. He, X. Li, M. Wang, G. Li, and L. Lin, “Spectral data quality assessment based on variability analysis: application to noninvasive hemoglobin measurement by dynamic spectrum,” *Analytical Methods*, vol. 7, no. 13, pp. 5565–5573, 2015.
- [108] M. Kumar, A. Veeraraghavan, and A. Sabharwal, “Distanceppg: Robust non-contact vital signs monitoring using a camera,” *Biomedical optics express*, vol. 6, no. 5, pp. 1565–1588, 2015.
- [109] C. Yu, Z. Liu, T. McKenna, A. T. Reisner, and J. Reifman, “A method for automatic identification of reliable heart rates calculated from ecg and ppg waveforms,” *Journal of the American Medical Informatics Association*, vol. 13, no. 3, pp. 309–320, 2006.
- [110] S. Vadrevu and M. S. Manikandan, “Real-time ppg signal quality assessment system for improving battery life and false alarms,” *IEEE transactions on circuits and systems II: express briefs*, vol. 66, no. 11, pp. 1910–1914, 2019.
- [111] C.-H. Goh, L. K. Tan, N. H. Lovell, S.-C. Ng, M. P. Tan, and E. Lim, “Robust ppg motion artifact detection using a 1-d convolution neural network,” *Computer methods and programs in biomedicine*, vol. 196, p. 105596, 2020.

- [112] B. L. Hill, N. Rakocz, Á. Rudas, J. N. Chiang, S. Wang, I. Hofer, M. Cannesson, and E. Halperin, "Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.
- [113] T. Aydemir, M. Sahin, and O. Aydemir, "Determination of hypertension disease using chirp z-transform and statistical features of optimal bandpass filtered short-time photoplethysmography signals," *Biomedical Physics & Engineering Express*, vol. 6, no. 6, p. 065033, 2020.
- [114] B. Huang, W. Chen, C.-L. Lin, C.-F. Juang, and J. Wang, "Mlp-bp: A novel framework for cuffless blood pressure measurement with ppg and ecg signals based on mlp-mixer neural networks," *Biomedical Signal Processing and Control*, vol. 73, p. 103404, 2022.
- [115] P. H. Charlton, B. Paliakaité, K. Pilt, M. Bachler, S. Zanelli, D. Kulin, J. Allen, M. Hallab, E. Bianchini, C. C. Mayer *et al.*, "Assessing hemodynamics from the photoplethysmogram to gain insights into vascular age: A review from vascagenet," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 322, no. 4, pp. H493–H522, 2022.
- [116] R. Couceiro, P. Carvalho, R. Paiva, J. Henriques, I. Quintal, M. Antunes, J. Muchlsteff, C. Eickholt, C. Brinkmeyer, M. Kelm *et al.*, "Assessment of cardiovascular function from multi-gaussian fitting of a finger photoplethysmogram," *Physiological measurement*, vol. 36, no. 9, p. 1801, 2015.
- [117] B. Smith and V. Madigan, "Understanding the haemodynamics of hypertension," *Current Hypertension Reports*, vol. 20, pp. 1–12, 2018.
- [118] M. O'rourke, "Vascular impedance in studies of arterial and cardiac function." *Physiological reviews*, vol. 62, no. 2, pp. 570–623, 1982.
- [119] B. Saugel, K. Kouz, T. W. Scheeren, G. Grewe, P. Hoppe, S. Romagnoli, and D. de Backer, "Cardiac output estimation using pulse wave analysis—physiology, algorithms, and technologies: A narrative review," *British journal of anaesthesia*, vol. 126, no. 1, pp. 67–76, 2021.
- [120] C. Liu, D. Zheng, A. Murray, and C. Liu, "Modeling carotid and radial artery pulse pressure waveforms by curve fitting with gaussian functions," *Biomedical Signal Processing and Control*, vol. 8, no. 5, pp. 449–454, 2013.
- [121] L. Wang, L. Xu, S. Feng, M. Q.-H. Meng, and K. Wang, "Multi-gaussian fitting for pulse waveform using weighted least squares and multi-criteria decision making method," *Computers in biology and medicine*, vol. 43, no. 11, pp. 1661–1672, 2013.
- [122] U. Rubins, "Finger and ear photoplethysmogram waveform analysis by fitting with gaussians," *Medical & biological engineering & computing*, vol. 46, pp. 1271–1276, 2008.
- [123] T. Nagasawa, K. Iuchi, R. Takahashi, M. Tsunomura, R. P. de Souza, K. Ogawa-Ochiai, N. Tsumura, and G. C. Cardoso, "Blood pressure estimation by photoplethysmogram decomposition into hyperbolic secant waves," *Applied Sciences*, vol. 12, no. 4, p. 1798, 2022.
- [124] K. Takazawa, N. Tanaka, M. Fujita, O. Matsuoka, T. Saiki, M. Aikawa, S. Tamura, and C. Ibukiyama, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform," *Hypertension*, vol. 32, no. 2, pp. 365–370, 1998.
- [125] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, "Detection of a and b waves in the acceleration photoplethysmogram," *Biomedical engineering online*, vol. 13, no. 1, pp. 1–18, 2014.
- [126] M. Elgendi, "Detection of c, d, and e waves in the acceleration photoplethysmogram," *Computer methods and programs in biomedicine*, vol. 117, no. 2, pp. 125–136, 2014.
- [127] M. J. Bourgeois, B. K. Gilbert, D. E. Donald, and E. H. Wood, "Characteristics of aortic diastolic pressure decay with application to the continuous monitoring of changes in peripheral vascular resistance," *Circulation research*, vol. 35, no. 1, pp. 56–66, 1974.
- [128] S. A. Esper and M. R. Pinsky, "Arterial waveform analysis," *Best Practice & Research Clinical Anaesthesiology*, vol. 28, no. 4, pp. 363–380, 2014.
- [129] R. Mukkamala, A. T. Reisner, H. M. Hojman, R. G. Mark, and R. J. Cohen, "Continuous cardiac output monitoring by peripheral blood pressure waveform analysis," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 459–467, 2006.
- [130] J.-H. Jeong, B. Lee, J. Hong, T.-H. Yang, and Y.-H. Park, "Reproduction of human blood pressure waveform using physiology-based cardiovascular simulator," *Scientific Reports*, vol. 13, no. 1, p. 7856, 2023.
- [131] P. S. Addison, "Slope transit time (stt): A pulse transit time proxy requiring only a single signal fiducial point," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2441–2444, 2016.
- [132] E. C.-P. Chua, S. J. Redmond, G. McDarby, and C. Heneghan, "Towards using photo-plethysmogram amplitude to measure blood pressure during sleep," *Annals of biomedical engineering*, vol. 38, no. 3, pp. 945–954, 2010.
- [133] K. Natarajan, R. C. Block, M. Yavarimanesh, A. Chandrasekhar, L. K. Mestha, O. T. Inan, J.-O. Hahn, and R. Mukkamala, "Photoplethysmography fast upstroke time intervals can be useful features for cuff-less measurement of blood pressure changes in humans," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 1, pp. 53–62, 2021.
- [134] X. Teng and Y. Zhang, "Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, vol. 4. IEEE, 2003, pp. 3153–3156.
- [135] S. C. Millasseau, R. Kelly, J. Ritter, and P. Chowienczyk, "Determination of age-related increases in large artery stiffness by digital pulse contour analysis," *Clinical science*, vol. 103, no. 4, pp. 371–377, 2002.
- [136] D. Fujita, A. Suzuki, and K. Ryu, "Ppg-based systolic blood pressure estimation method using pls and level-crossing feature," *Applied Sciences*, vol. 9, no. 2, p. 304, 2019.
- [137] E. von Wowern, G. Östling, P. M. Nilsson, and P. Olofsson, "Digital photoplethysmography for assessment of arterial stiffness: repeatability and comparison with applanation tonometry," *PloS one*, vol. 10, no. 8, p. e0135659, 2015.
- [138] S. Ghosh, B. P. Chattopadhyay, R. M. Roy, J. Mukherjee, and M. Mahadevappa, "Non-invasive cuffless blood pressure and heart rate monitoring using impedance cardiography," *Intelligent Medicine*, 2021.
- [139] S. Haddad, A. Boukhayma, and A. Caizzone, "Continuous ppg-based blood pressure monitoring using multi-linear regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2096–2105, 2021.
- [140] D. Park, S. J. Cho, K. Kim, H. Woo, J. E. Kim, J.-Y. Lee, J. Koh, J. Lee, J. S. Choi, D. K. Chang *et al.*, "Prediction algorithms for blood pressure based on pulse wave velocity using health checkup data in healthy korean men: Algorithm development and validation," *JMIR medical informatics*, vol. 9, no. 12, p. e29212, 2021.
- [141] R. Mieloszyk, H. Twede, J. Lester, J. Wander, S. Basu, G. Cohn, G. Smith, D. Morris, S. Gupta, D. Tan *et al.*, "A comparison of wearable tonometry, photoplethysmography, and electrocardiography for cuffless measurement of blood pressure in an ambulatory setting," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 2864–2875, 2022.
- [142] R. Arathy, P. Nabeel, J. Joseph, and M. Sivaprakasam, "Accelerometric patch probe for cuffless blood pressure evaluation from carotid local pulse wave velocity: Design, development, and in vivo experimental study," *Biomedical Physics & Engineering Express*, vol. 5, no. 4, p. 045010, 2019.
- [143] A. Dagamseh, Q. Qananwah, H. Al Quran, and K. S. Ibrahim, "Towards a portable-noninvasive blood pressure monitoring system utilizing the photoplethysmogram signal," *Biomedical Optics Express*, vol. 12, no. 12, pp. 7732–7751, 2021.
- [144] N. Jean Effil and R. Rajeswari, "Wavelet scattering transform and long short-term memory network-based noninvasive blood pressure estimation from photoplethysmograph signals," *Signal, Image and Video Processing*, vol. 16, no. 1, pp. 1–9, 2022.
- [145] M. W. K. Fong, E. Ng, K. E. Z. Jian, and T. J. Hong, "Svr ensemble-based continuous blood pressure prediction using multi-channel photoplethysmogram," *Computers in biology and medicine*, vol. 113, p. 103392, 2019.
- [146] N. Hasanzadeh, M. M. Ahmadi, and H. Mohammadzade, "Blood pressure estimation using photoplethysmogram signal and its morphological features," *IEEE Sensors Journal*, vol. 20, no. 8, pp. 4300–4310, 2019.
- [147] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 859–869, 2016.
- [148] J. Zhang, D. Wu, and Y. Li, "Cuff-less and calibration-free blood pressure estimation using convolutional autoencoder with unsupervised feature extraction," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 3323–3326.
- [149] Z. Liu, F. Miao, R. Wang, J. Liu, B. Wen, and Y. Li, "Cuff-less blood pressure measurement based on deep convolutional neural network," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 3775–3778.
- [150] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network," *Sensors*, vol. 19, no. 15, p. 3420, 2019.

- [151] D. U. Jeong and K. M. Lim, "Combined deep cnn-lstm network-based multitasking learning architecture for noninvasive continuous blood pressure estimation using difference in ecg-ppg features," *Scientific Reports*, vol. 11, no. 1, pp. 1–8, 2021.
- [152] C. Ma, P. Zhang, F. Song, Y. Sun, G. Fan, T. Zhang, Y. Feng, and G. Zhang, "Kd-informer: cuff-less continuous blood pressure waveform estimation approach based on single photoplethysmography," *IEEE Journal of Biomedical and Health Informatics*, 2022.
- [153] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.
- [154] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [155] A. D. Choudhury, R. Banerjee, A. Sinha, and S. Kundu, "Estimating blood pressure using windkessel model on photoplethysmogram," in *2014 36th annual international conference of the IEEE engineering in medicine and biology society*. IEEE, 2014, pp. 4567–4570.
- [156] S. Datta, R. Banerjee, A. D. Choudhury, A. Sinha, and A. Pal, "Blood pressure estimation from photoplethysmogram using latent parameters," in *2016 IEEE International conference on communications (ICC)*. IEEE, 2016, pp. 1–7.
- [157] B. Jana, K. Oswal, S. Mitra, G. Saha, and S. Banerjee, "Windkessel model-based cuffless blood pressure estimation using continuous wave doppler ultrasound system," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 9989–9999, 2020.
- [158] K. A. Beeks, "Arterial blood pressure estimation using ultrasound technology and transmission line arterial model," Ph.D. dissertation, Massachusetts Institute of Technology, 2019.
- [159] C. Poon and Y. Zhang, "Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time," in *2005 IEEE engineering in medicine and biology 27th annual conference*. IEEE, 2006, pp. 5877–5880.
- [160] C. M. Quick, W. L. Young, and A. Noordergraaf, "Infinite number of solutions to the hemodynamic inverse problem," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 280, no. 4, pp. H1472–H1479, 2001.
- [161] C. M. Quick, D. S. Berger, R. H. Stewart, G. A. Laine, C. J. Hartley, and A. Noordergraaf, "Resolving the hemodynamic inverse problem," *IEEE transactions on biomedical engineering*, vol. 53, no. 3, pp. 361–368, 2006.
- [162] F. Schrumpf, P. Frenzel, C. Aust, G. Osterhoff, and M. Fuchs, "Assessment of non-invasive blood pressure prediction from ppg and rppg signals using deep learning," *Sensors*, vol. 21, no. 18, p. 6022, 2021.
- [163] Y. Yang, K. Zha, Y. Chen, H. Wang, and D. Katabi, "Delving into deep imbalanced regression," in *International Conference on Machine Learning*. PMLR, 2021, pp. 11842–11851.
- [164] L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, "Smote for regression," in *Portuguese conference on artificial intelligence*. Springer, 2013, pp. 378–389.
- [165] E. O'Brien, J. Petrie, W. Littler, M. De Swiet, P. L. Padfield, D. Altman, M. Bland, A. Coats, N. Atkins *et al.*, "The british hypertension society protocol for the evaluation of blood pressure measuring devices," *J hypertens*, vol. 11, no. Suppl 2, pp. S43–S62, 1993.
- [166] I. S. Association *et al.*, "Ieee standard for wearable cuffless blood pressure measuring devices," *IEEE Std*, pp. 1708–2014, 2014.
- [167] E. O'Brien, N. Atkins, G. Stergiou, N. Karpettas, G. Parati, R. Asmar, Y. Imai, J. Wang, T. Mengden, A. Shennan *et al.*, "European society of hypertension international protocol revision 2010 for the validation of blood pressure measuring devices in adults," *Blood pressure monitoring*, vol. 15, no. 1, pp. 23–38, 2010.
- [168] G. S. Stergiou, A. P. Avolio, P. Palatini, K. G. Kyriakoulis, A. E. Schutte, S. Mieke, A. Kollias, G. Parati, R. Asmar, N. Pantazis *et al.*, "European society of hypertension recommendations for the validation of cuffless blood pressure measuring devices: European society of hypertension working group on blood pressure monitoring and cardiovascular variability," *Journal of Hypertension*, pp. 10–1097, 2023.
- [169] G. S. Stergiou, B. Alpert, S. Mieke, R. Asmar, N. Atkins, S. Eckert, G. Frick, B. Friedman, T. Graßl, T. Ichikawa *et al.*, "A universal standard for the validation of blood pressure measuring devices: Association for the advancement of medical instrumentation/european society of hypertension/international organization for standardization (aami/esh/iso) collaboration statement," *Hypertension*, vol. 71, no. 3, pp. 368–374, 2018.
- [170] "Non-Invasive Sphygmomanometers - Part 3: Clinical Investigation Of Continuous Automated Measurement Type," American National Standard, Standard, Dec. 2022.
- [171] R. Mukkamala and J.-O. Hahn, "Toward ubiquitous blood pressure monitoring via pulse transit time: Predictions on maximum calibration period and acceptable error limits," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1410–1420, 2017.
- [172] A.-L. Boulesteix, V. Stierle, and A. Hapfelmeier, "Publication bias in methodological computational research," *Cancer informatics*, vol. 14, pp. CIN-S30 747, 2015.
- [173] J. P. Ioannidis, "Why most published research findings are false," *PLoS medicine*, vol. 2, no. 8, p. e124, 2005.
- [174] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.
- [175] Y. Liang, Z. Chen, G. Liu, and M. Elgendi, "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in china," *Scientific data*, vol. 5, no. 1, pp. 1–7, 2018.
- [176] H.-C. Lee, Y. Park, S. B. Yoon, S. M. Yang, D. Park, and C.-W. Jung, "Vitaldb, a high-fidelity multi-parameter vital signs database in surgical patients," *Scientific Data*, vol. 9, no. 1, pp. 1–9, 2022.
- [177] Z.-D. Liu, J.-K. Liu, B. Wen, Q.-Y. He, Y. Li, and F. Miao, "Cuffless blood pressure estimation using pressure pulse wave signals," *Sensors*, vol. 18, no. 12, p. 4227, 2018.
- [178] Q. Wu, J. Yang, G. Zheng, X. Zhang, Y. Zhang, C. Xu, and J. Tang, "An ambulatory blood pressure monitoring system based on the uncalibrated steps of the wrist," in *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2019, pp. 1–6.
- [179] A. Abolhasani, M. Mousazadeh, and A. Khoei, "Real-time, cuff-less and non-invasive blood pressure monitoring," *Informacije MIDEM*, vol. 50, no. 2, pp. 87–104, 2020.
- [180] D. Nachman, Y. Gepner, N. Goldstein, E. Kabakov, A. B. Ishay, R. Littman, Y. Azmon, E. Jaffe, and A. Eisenkraft, "Comparing blood pressure measurements between a photoplethysmography-based and a standard cuff-based manometry device," *Scientific reports*, vol. 10, no. 1, pp. 1–9, 2020.
- [181] D. Nachman, A. Gilan, N. Goldstein, K. Constantini, R. Littman, A. Eisenkraft, E. Grossman, and Y. Gepner, "Twenty-four-hour ambulatory blood pressure measurement using a novel noninvasive, cuffless, wireless device," *American Journal of Hypertension*, vol. 34, no. 11, pp. 1171–1180, 2021.
- [182] E. Kachel, K. Constantini, D. Nachman, S. Carasso, R. Littman, A. Eisenkraft, and Y. Gepner, "A pilot study of blood pressure monitoring after cardiac surgery using a wearable, non-invasive sensor," *Frontiers in Medicine*, p. 1284, 2021.
- [183] P. Schoettker, J. Degott, G. Hofmann, M. Proen  a, G. Bonnier, A. Lemkaddem, M. Lemay, R. Schorer, U. Christen, J.-F. Knebel *et al.*, "Blood pressure measurements with the optibp smartphone app validated against reference auscultatory measurements," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [184] J. Degott, A. Ghajarzadeh-Wurzner, G. Hofmann, M. Proen  a, G. Bonnier, A. Lemkaddem, M. Lemay, U. Christen, J.-F. Knebel, V. Durgnat *et al.*, "Smartphone based blood pressure measurement: accuracy of the optibp mobile application according to the aami/esh/iso universal validation protocol," *Blood Pressure Monitoring*, vol. 26, no. 6, p. 441, 2021.
- [185] O. Desebbe, C. Anas, B. Alexander, K. Kouz, J.-F. Knebel, P. Schoettker, J. Creteur, J.-L. Vincent, and A. Joosten, "Evaluation of a novel optical smartphone blood pressure application: A method comparison study against invasive arterial blood pressure monitoring in intensive care unit patients," 2022.
- [186] G. Hofmann, M. Proen  a, J. Degott, G. Bonnier, A. Lemkaddem, M. Lemay, R. Schorer, U. Christen, J.-F. Knebel, and P. Schoettker, "A novel smartphone app for blood pressure measurement: a proof-of-concept study against an arterial catheter," *Journal of Clinical Monitoring and Computing*, pp. 1–11, 2022.
- [187] M. Caillat, J. Degott, A. Wuerzner, M. Proen  a, G. Bonnier, J.-F. Knebel, C. Stoll, U. Christen, V. Durgnat, G. Hofmann *et al.*, "Accuracy of blood pressure measurement across bmi categories using the optibpTM mobile application," *Blood Pressure*, vol. 31, no. 1, pp. 288–296, 2022.
- [188] O. Desebbe, A. Tighenifi, A. Jacobs, L. Toubal, Y. Zekhini, D. Chirnoaga, V. Collange, B. Alexander, J. F. Knebel, P. Schoettker *et al.*, "Evaluation of a novel mobile phone application for blood pressure monitoring: a proof of concept study," *Journal of Clinical Monitoring and Computing*, vol. 36, no. 4, pp. 1147–1153, 2022.
- [189] J. Sola, M. Bertschi, and J. Krauss, "Measuring pressure: Introducing obpm, the optical revolution for blood pressure monitoring," *IEEE pulse*, vol. 9, no. 5, pp. 31–33, 2018.

- [190] J. Solà, Y. Ghamri, M. Proençá, F. Braun, A. Lemkadem, N. Pierrel, C. Verjus, M. Bertschi, and P. Schoettker, "Tracking blood pressure changes in anesthetized patients: the optical blood pressure monitoring (obpm) technology," in *Conference Paper 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, At Honolulu, Honolulu, Hawaii, USA*, 2018.
- [191] J. Solà, A. Vybornova, S. Fallet, E. Olivero, B. De Marco, O. Grossenbacher, N. Ignjatovic, B. Ignjatovic, M. Favre-Bulle, N. Levinson *et al.*, "Are cuffless devices challenged enough? design of a validation protocol for ambulatory blood pressure monitors at the wrist: the case of the aktiia bracelet," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 4437–4440.
- [192] C. Pellaton, A. Vybornova, S. Fallet, L. Marques, O. Grossenbacher, B. De Marco, V. Chapuis, M. Bertschi, B. S. Alpert, and J. Solà, "Accuracy testing of a new optical device for noninvasive estimation of systolic and diastolic blood pressure compared to intra-arterial measurements," *Blood Pressure Monitoring*, vol. 25, no. 2, pp. 105–109, 2020.
- [193] A. Vybornova, E. Polychronopoulou, A. Wurzner-Ghajarzadeh, S. Fallet, J. Sola, and G. Wuerzner, "Blood pressure from the optical aktiia bracelet: a 1-month validation study using an extended iso81060-2 protocol adapted for a cuffless wrist device," *Blood Pressure Monitoring*, vol. 26, no. 4, p. 305, 2021.
- [194] J. Sola, A. Vybornova, S. Fallet, E. Polychronopoulou, A. Wurzner-Ghajarzadeh, and G. Wuerzner, "Validation of the optical aktiia bracelet in different body positions for the persistent monitoring of blood pressure," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [195] J. Sola, M. Cortes, D. Perruchoud, B. De Marco, M. D. Lobo, C. Pellaton, G. Wuerzner, N. D. Fisher, and J. Shah, "Guidance for the interpretation of continual cuffless blood pressure data for the diagnosis and management of hypertension," *Frontiers in Medical Technology*, p. 25, 2022.
- [196] J. Alexandre, K. Tan, T. P. Almeida, J. Sola, B. S. Alpert, and J. Shah, "Validation of the aktiia blood pressure cuff for clinical use according to the ansi/aami/iso 81060-2: 2013 protocol," *Blood Pressure Monitoring*, vol. 28, no. 2, p. 109, 2023.
- [197] T. P. Almeida, M. Cortés, D. Perruchoud, J. Alexandre, P. Vermare, J. Sola, J. Shah, L. Marques, and C. Pellaton, "Aktiia cuffless blood pressure monitor yields equivalent daytime blood pressure measurements compared to a 24-h ambulatory blood pressure monitor: Preliminary results from a prospective single-center study," *Hypertension Research*, pp. 1–6, 2023.
- [198] K.-G. Ng, C.-M. Ting, J.-H. Yeo, K.-W. Sim, W.-L. Peh, N.-H. Chua, N.-K. Chua, and F. Kwong, "Progress on the development of the mediwatch ambulatory blood pressure monitor and related devices," *Blood Pressure Monitoring*, vol. 9, no. 3, pp. 149–165, 2004.
- [199] D. Nair, S. Tan, H. Gan, S. Lim, J. Tan, M. Zhu, H. Gao, N. Chua, W. Peh, and K. Mak, "The use of ambulatory tonometric radial arterial wave capture to measure ambulatory blood pressure: the validation of a novel wrist-bound device in adults," *Journal of human hypertension*, vol. 22, no. 3, pp. 220–222, 2008.
- [200] S. Theilade, C. Joergensen, F. Persson, M. Lajer, and P. Rossing, "Ambulatory tonometric blood pressure measurements in patients with diabetes," *Diabetes technology & therapeutics*, vol. 14, no. 6, pp. 453–456, 2012.
- [201] L. Garcia-Ortiz, J. I. Recio-Rodríguez, J. J. Canales-Reina, A. Cabrejas-Sánchez, A. Gomez-Arranz, J. F. Magdalena-Belio, N. Guenaga-Saenz, C. Agudo-Conde, and M. A. Gomez-Marcos, "Comparison of two measuring instruments, b-pro and sphygmocor system as reference, to evaluate central systolic blood pressure and radial augmentation index," *Hypertension Research*, vol. 35, no. 6, pp. 617–623, 2012.
- [202] T. Komori, K. Eguchi, S. Hoshide, B. Williams, and K. Kario, "Comparison of wrist-type and arm-type 24-h blood pressure monitoring devices for ambulatory use," *Blood Pressure Monitoring*, vol. 18, no. 1, pp. 57–62, 2013.
- [203] T. J. Doty, B. Kelliher, T.-P. Jung, J. K. Zao, and I. Litvan, "The wearable multimodal monitoring system: A platform to study falls and near-falls in the real-world," in *International conference on human aspects of IT for the aged population*. Springer, 2015, pp. 412–422.
- [204] A. Jakes, J. Wade, Z. Vowles, P. T. Seed, A. H. Shennan, L. C. Chappell, and D. Nzelu, "Validation of the bpro radial pulse waveform acquisition device in pregnancy and gestational hypertensive disorders," *Blood Pressure Monitoring*, vol. 26, no. 5, pp. 380–384, 2021.
- [205] T. S. Shoot, M. Weenk, T. H. van de Belt, L. J. Engelen, H. van Goor, and S. J. Bredie, "A new cuffless device for measuring blood pressure: a real-life validation study," *Journal of medical Internet research*, vol. 18, no. 5, p. e5414, 2016.
- [206] P. A. Ogink, J. M. de Jong, M. Koeneman, M. Weenk, L. J. Engelen, H. van Goor, T. H. van de Belt, and S. J. Bredie, "Feasibility of a new cuffless device for ambulatory blood pressure measurement in patients with hypertension: mixed methods study," *Journal of Medical Internet Research*, vol. 21, no. 6, p. e11164, 2019.
- [207] P. A. Ogink, J. M. de Jong, M. Koeneman, M. Weenk, L. JLPG, H. v. G. Engelen, T. H. van de Belt, S. J. Bredie, and S. Bredie, "Usability of a new cuffless device for ambulatory blood pressure measurement in hypertension."
- [208] P. Krisai, A. S. Vischer, L. Kilian, A. Meienberg, M. Mayr, and T. Burkard, "Accuracy of 24-hour ambulatory blood pressure monitoring by a novel cuffless device in clinical practice," *Heart*, vol. 105, no. 5, pp. 399–405, 2019.
- [209] J. Nyvad, K. L. Christensen, N. H. Buus, and M. Reinhard, "The cuffless somnotouch nibrp device shows poor agreement with a validated oscillometric device during 24-h ambulatory blood pressure monitoring," *The Journal of Clinical Hypertension*, vol. 23, no. 1, pp. 61–70, 2021.
- [210] M. H. McGillion, N. Dvirnik, S. Yang, E. Belley-Côté, A. Lamy, R. Whitlock, M. Marcucci, F. K. Borges, E. Duceppe, C. Ouellette *et al.*, "Continuous noninvasive remote automated blood pressure monitoring with novel wearable technology: A preliminary validation study," *JMIR mHealth and uHealth*, vol. 10, no. 2, p. e24916, 2022.
- [211] G. Sayer, G. Piper, E. Vorovich, J. Raikhelkar, G. H. Kim, D. Rodgers, D. Shimbo, and N. Uriel, "Continuous monitoring of blood pressure using a wrist-worn cuffless device," *American Journal of Hypertension*, vol. 35, no. 5, pp. 407–413, 2022.
- [212] "Internally validated independent results for noninvasive, cuff-less blood pressure estimation utilizing valencell's deep-ppg technology," Valencell, Inc, Tech. Rep., 2020.
- [213] Y. Zhang, C. Zhou, Z. Huang, and X. Ye, "Study of cuffless blood pressure estimation method based on multiple physiological parameters," *Physiological Measurement*, vol. 42, no. 5, p. 055004, 2021.
- [214] A.-W. Harzing, "Publish or Perish," <https://harzing.com/resources/publish-or-perish>, 2007.