Features from the photoplethysmogram and the electrocardiogram for estimating changes in blood pressure

Eoin Finnegan[1,*]**, Shaun Davidson**[1]**, Mirae Harford**[1,2,3]**, Peter Watkinson**[2,3]**, Lionel Tarassenko**[1]**, and Mauricio Villarroel**[1]

[1]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK
[2]Critical Care Research Group, Nuffield Department of Clinical Neurosciences, University of Oxford
[3]NIHR Oxford Biomedical Research Centre, Oxford, UK.
[*]eoin.finnegan@eng.ox.ac.uk

ABSTRACT

There is increasing emphasis being placed on cuffless blood pressure (BP) monitoring using the electrocardiogram (ECG) and photoplethysmogram (PPG). These signals have previously been employed to compute estimates of the pulse arrival time (PAT, the time between characteristic fiducial points on the ECG and PPG) as a surrogate for BP. Current work in this field is focused on using either the ECG or the PPG alone as a single source BP estimator using characteristic morphological features and machine learning models. However, the appropriate features and models to use remain unclear. As a result, BP estimation using the ECG or the PPG signals alone has produced inconclusive results. In this work, we investigated the best features available from the PPG and the ECG for BP estimation using both linear and non-linear models.

We conducted a clinical study involving 30 healthy volunteers (53.8% female, 28 ($\pm$ 9) years old, with a body mass index of 22.5 $\pm$ (5.2 kg/m^2). Each session lasted 28.0 ($\pm$ 0.12) minutes and BP was varied by administering an infusion of phenylephrine, a medication that causes arterial and venous vasoconstriction. We extracted a large and diverse set of features from both the PPG and the ECG and assessed their individual importance for estimating changes in BP (ΔBP) using a ranking coefficient. In addition to features commonly used in the literature, we propose new features extracted from both signals. We implemented linear (ordinary least squares, OLS) and non-linear (random forest, RF) machine learning models to estimate ΔBP. We adopted a *hybrid calibration* strategy by including patient demographics in the feature set. We trained, tuned, and evaluated these models in a nested leave-one-subject-out cross-validation framework and we reported the results as correlation coefficient (ρ_p), root mean squared error (RMSE), and mean absolute error (MAE). We compared our results to those of estimating ΔBP using PAT.

The non-linear RF model significantly ($p < 0.05$) outperformed the linear OLS model using both the PPG and the ECG signals across all performance metrics. Estimating ΔSBP using the PPG alone (ρ_p = 0.86 (0.23), RMSE = 5.66 (4.76) mmHg, MAE = 4.86 (4.29) mmHg) performed significantly better than using the ECG alone (ρ_p = 0.69 (0.45), RMSE = 6.79 (4.76) mmHg, MAE = 5.28 (4.57) mmHg), all $p < 0.001$. Estimating ΔBP using features from the PPG alone had a similar performance to that of using PAT (which requires a simultaneous ECG signal). Kurtosis of the PPG waveform showed consistently high feature ranking for both the OLS and RF models. Additionally, the highest ranking features from the PPG largely modelled increasing reflected wave interference driven by changes in arterial stiffness. This finding was supported by changes observed in the PPG waveform in response to the phenylephrine infusion. However, a large number of features were required for accurate BP estimation, highlighting the high complexity of the problem.

We conclude that the PPG alone may be further explored as a potential single source, cuffless, blood pressure estimator. The use of the ECG alone is not justified. Non-linear models may perform better as they are able to incorporate interactions between feature values and demographics. However, demographics may not adequately account for the unique and individualised relationship between the extracted features and BP.

1 Introduction

Changes in the cardiovascular and autonomic nervous systems are reflected in changes in signals such as the photoplethysmogram (PPG) and the electrocardiogram (ECG). These physiological signals are ubiquitous in a clinical setting and increasingly in an out-of-clinic setting due to the development of wearables such as smartwatches. As a result, recent advances in non-invasive, cuffless estimation of blood pressure (BP) have been focused on utilising the PPG and the ECG signals. For example, pulse arrival time (PAT), computed as the time difference between two fiducial points in the ECG and PPG waveforms, has been shown in certain studies to have strong correlations with BP[1]. However, measuring PAT requires two synchronous devices, both

of which are susceptible to independent sources of noise such as motion artefacts. Additionally, factors such as the pre-ejection period (PEP) (the time delay between the electrical depolarisation of the heart's left ventricle and the opening of the aortic valve) may impact PAT estimates independently of BP[2]. A summary of BP estimation methods using PAT can be found in[3–5].

Driven by modern techniques, research on cuffless BP estimation has increasingly focused on relating morphological features of the PPG and the ECG waveforms to BP (or changes in BP, ΔBP) using data-driven models. However, the optimal features and models required for accurate BP estimation remain unclear. In this paper, we implemented linear and non-linear machine learning (ML) models to estimate ΔBP using a large and diverse cohort of features from the PPG and the ECG waveforms. In addition to features commonly used in the literature, we proposed new features from both signals and assessed their individual importance for estimating ΔBP. We compared our results to those estimating changes in BP using PAT, and evaluated the PPG and ECG as potential single source devices for BP estimation. This work was carried out using data from a clinical study involving 30 healthy volunteers. Changes in BP were induced by the administration of phenylephrine, a vasoactive medication that causes arterial and venous vasoconstriction and increases cardiac preload (initial stretching of cardiac muscles).

1.1 Relationship between changes in BP and changes in the PPG

The pulsatile PPG waveform is related to changes in blood volume over time in a bed of tissue[6]. The PPG signal is typically recorded by a pulse oximeter placed on the index finger. Methods using the PPG have been proposed as a continuous, non-invasive, cuffless approach to estimate BP[7]. The PPG-BP relationship is driven, in part, by the theoretical relationship between changes in pressure and volume of blood in a localised region of the arteries[7], as well as the impact of reflected waves[8]. The PPG offers significant benefits for BP monitoring over conventional cuff-based measurements. Most notably, the PPG can be recorded by a single, unobtrusive optical sensor which also has the potential to be implemented on a wearable device such as a smartwatch[9,10]. However, there is no generally accepted method relating changes in the PPG waveform to ΔBP and, as a result, a variety of different approaches are proposed in the literature. A summary of BP estimation algorithms using the PPG waveform can be found in[7,11,12].

Much like PAT[1], features extracted from the PPG waveform are thought to be subject-specific requiring individual calibration for accurate mapping to BP values[7]. Mukkamala *et al.*[7] splits calibration strategies into three groups: *individual*, *hybrid* and *population*. In *individual calibration*, all model parameters are determined by multiple paired recordings of BP and PPG from a single individual. This approach may be feasible for BP estimation using PAT (where typically only two parameters are required for modelling[1]), however it becomes intractable for multi-parameter ML models often employed for PPG-based BP estimation. In *hybrid calibration*, only one calibration BP data point is required for a single individual. The remaining model parameters are estimated using the individual's demographics and a training set comprised of multiple BP-PPG pairs from a cohort of different individuals. In *population calibration*, no calibration recordings are required and all calibration is handled using the individual's demographics and a similar training set. Certain dependencies on the morphology of the PPG waveform have been previously reported for age[13], sex[14] and body mass index (BMI)[15], however these are often not strong enough to allow for good accuracy when using *population calibration* strategies. As a result, the majority of studies proposing PPG-based BP estimation opt for a *hybrid calibration* strategy.

Sun *et al.*[16] evaluated the use of a linear regression model for estimating systolic blood pressure (SBP) measured by a commercial Portapres device (Finapres), using the volume-clamp method. Nineteen subjects underwent an exercise test followed by a posture change test. Nineteen features were extracted from the PPG. Using a *hybrid calibration* strategy, the authors reported a root mean squared error (RMSE) of 8.99 mmHg and Pearson's correlation coefficient (ρ_p) of 0.85 during the exercise test, and a RMSE of 7.33 mmHg and ρ_p of 0.47 during the posture change test. Normalised weights of the linear model were used to highlight features that had the most predictive power and an inconsistency of the best features was found between the two tests. Miao *et al.*[17] used support vector regression on 14 extracted features from 73 subjects to track BP changes induced by physical exercise. A genetic algorithm for feature selection was developed to highlight features that best estimated ΔBP. The stability of the proposed models was evaluated in a follow-up study for 1 day, 10 days and 6 months after the initial test. The results suggested that, similar to PAT[18,19], the models lose their accuracy over time and therefore require frequent recalibration. Hasanzadeh *et al.*[20] extracted features from the PPG in a subset of 1,000 individuals from the MIMIC-III dataset. The intra-arterial blood pressure was used as a reference. The authors implemented a linear regression model as well as a non-linear tree-based model, AdaBoost. AdaBoost significantly outperformed the linear model for their dataset. Additionally, the authors highlighted the sensitivity of PPG feature detection in the presence of random noise. Recent shifts in the field have moved towards the application of deep learning algorithms[21,22] which often do not require the extraction of handcrafted features and instead work on the raw PPG waveform. However, while these approaches are reported to improve estimation accuracy, they lose out in model interpretability due to their black box nature.

1.2 Relationship between changes in BP and changes in the ECG

The ECG is a measure of the electrical activity generated in the myocardium (heart muscle) during each heartbeat. It is acquired by measuring the voltage difference between two points on the body surface over time[23]. A single lead ECG is typically

recorded by three electrodes placed on an individual's torso forming Einthoven's triangle[23].

In comparison to the PPG, less focus has been placed on the potential use of the ECG for BP estimation. The general theory governing the relationship between ECG and BP is based on a cyclical process known as mechano-electric coupling (MEC)[24]. Changes in the electrical properties of the heart have a direct impact on the contractility of the heart. This is known as excitation-contraction coupling. Similarly, changes in the mechanical properties of the tissues surrounding the heart are detected by mechanosensitive ion channels, resulting in local changes in the electrical potential. This is known as mechano-electrical feedback. MEC, therefore, describes the cyclical process whereby changes in the ECG waveform can reflect ΔBP. However, MEC is influenced not only by extra-cardiac control mechanisms such as the Autonomic Nervous System (ANS) and hormonal changes, but also by environmental mechanisms such as ion concentrations and temperature[25]. Therefore, the relationship between ΔBP that can be detected by analysing morphological changes in the ECG waveform are not as developed as methods using the PPG waveform.

Simjanoska *et al.*[26] used a dataset containing 51 individuals from a mixture of 4 open-source datasets. Three of the datasets (totalling 44 individuals) included healthy volunteers using commercial ECG sensors with reference BP values measured by a cuff. In these datasets BP was perturbed by natural variations with each individual contributing a range from 1 to 8 measurements overall. The fourth dataset was recorded from 7 patients with traumatic brain injuries in an intensive care unit (ICU) using clinical-grade ECG sensors and reference BP values measured by an arterial line. Complexity features such as mobility and entropy were extracted from 30-second segments of ECG data and used to train a random forest regression model using a train-validation-test split. After calibration, the model achieved mean absolute error (MAE) of 7.72 mmHg for SBP, 9.45 mmHg for diastolic blood pressure (DBP), and 8.13 mmHg for mean arterial blood pressure (MAP). However, the authors only implemented a small number of complexity features, and the mixture of data sources (especially with regard to mixing healthy and unhealthy populations) makes interpreting the results difficult.

2 Methods

2.1 Clinical study

Thirty healthy volunteers with no history of cardiovascular disease, were recruited for the clinical study. The study protocol has been outlined previously[1,27]. The study took place at the Cardiovascular Clinical Research Facility within the John Radcliffe Hospital, Oxford, UK. This study was reviewed and approved by the Oxford University Research and Ethics Committee and Clinical Trials and Research Governance teams (R63796/RE001). All methods were performed in accordance with the relevant guidelines and regulations. Individual informed written consent was obtained from all the participants in the study to record the data and publish the results.

The study was split into four main stages: rest, dose increase, maximum infusion, and washout. Each session began with a 5-minute resting period. Participants were then administered an infusion of phenylephrine. Phenylephrine is an α-1 adrenergic receptor agonist which causes arterial and venous vasoconstriction as well as reflex bradycardia[28]. The dosing regimen was guided by the clinician in the Medical Sciences Division Ethics Committee (University of Oxford) who balanced the desired clinical effect against any safety concerns[27]. Phenylephrine was administered as an intravenous solution, starting at a rate of 0.2mcg/kg/min with an increase of 0.2mcg/kg/min every 1 minute for 10 increments. Once the maximum rate was reached, the infusion and all monitoring remained constant for a further 6 minutes. Each session ended with an 8-minute washout period. Participants were asked to refrain from ingesting caffeinated drinks for four hours prior to the study visit as caffeine is a vasoconstrictor. Participants were in a supine position with the head and trunk raised to between 15 and 45 degrees (semi-Fowler's position).

Instrumentation

A Philips Intellivue MX800 patient monitor (Philips, Netherlands) recorded BP via a sphygmomanometer cuff. The Philips comfort care M3001A BP cuff was wrapped around the upper left arm of the participant with the centre directly above the brachial artery as recommended[29]. The ixTrend software (Ixellence GmbH, Germany) was used to record the data generated by the patient monitor. A Stowood Visi Black shadow polygraphy device (Stowood, UK) recorded the ECG and PPG signals. Three ECG electrodes were attached to the participant's torso forming Einthoven's triangle. A pulse oximeter (Masimo, USA) probe was placed on the participant's right hand. Table 1 provides a summary of the physiological parameters recorded by the Philips and Stowood monitors.

2.2 Features from the PPG

To preprocess the PPG waveform, we performed the following steps. The PPG signal was first filtered using an 8th-order Butterworth infinite impulse response (IIR) band-pass filter with cut-off frequencies of 0.5 Hz and 10 Hz as recommended in[3]. We then followed the work of Villarroel *et al.*.[30] to detect the pulse onsets and to assess the quality of the PPG signal (SQI$_{PPG}$). Amplitude modulation of the PPG can result from respiration[31] as well as changes in contact pressure of the sensor on the

Table 1. Physiological parameters and sampling rates recorded by the two devices used in this study.

Device	Name	Description	Sampling rate (Hz)
Philips	SBP, MAP, DBP	Blood pressure signal from the sphygmomanometer cuff	1/60
Stowood	ECG	1-lead electrocardiography signal	256
	PPG	Photoplethysmography signal	512
	HR	Heart rate derived from Masimo pulse oximeter	5

surface of the skin[32]. In order to reduce the dependence of this amplitude modulation as well as heart rate, we normalised the PPG beats to have unit amplitude and time duration. The first and second derivatives of the PPG beats, known as the velocity plethysmogram (VPG) and acceleration plethysmogram (APG) respectively, were extracted using a 7th-order Savitzky-Golay filter[33]. We used the nomenclature defined by Elgendi *et al.*.[34] for the characteristic fiducial points of the PPG, VPG and APG.

Figure 1 shows the typical changes in the morphology of the PPG, VPG and APG beat seen for a participant during the four main stages of the study protocol (rest, dose increase, maximum infusion and washout). The following fiducial points were detected using the criteria defined in[35]: dicrotic notch (N), diastolic peak (D) and the a, b, c, d, and e waves on the APG. The dicrotic notch defines the end of systole and the start of diastole. The systolic peak (S) was defined as the first turning point of the PPG pulse above its midpoint.

The features we extracted from the PPG are summarised in table 2. Features from the PPG are broken into five main categories: PPG morphology, VPG morphology, APG morphology, Gaussian decomposition and principal component analysis (PCA). The methodology for extracting features from the PPG morphology, VPG morphology and APG morphology are extensively documented in the literature[11,12] and the reader is directed to the relevant publications provided in table 2 for more information. The methodology for computing the Gaussian decomposition and PCA features are described in the sections below.
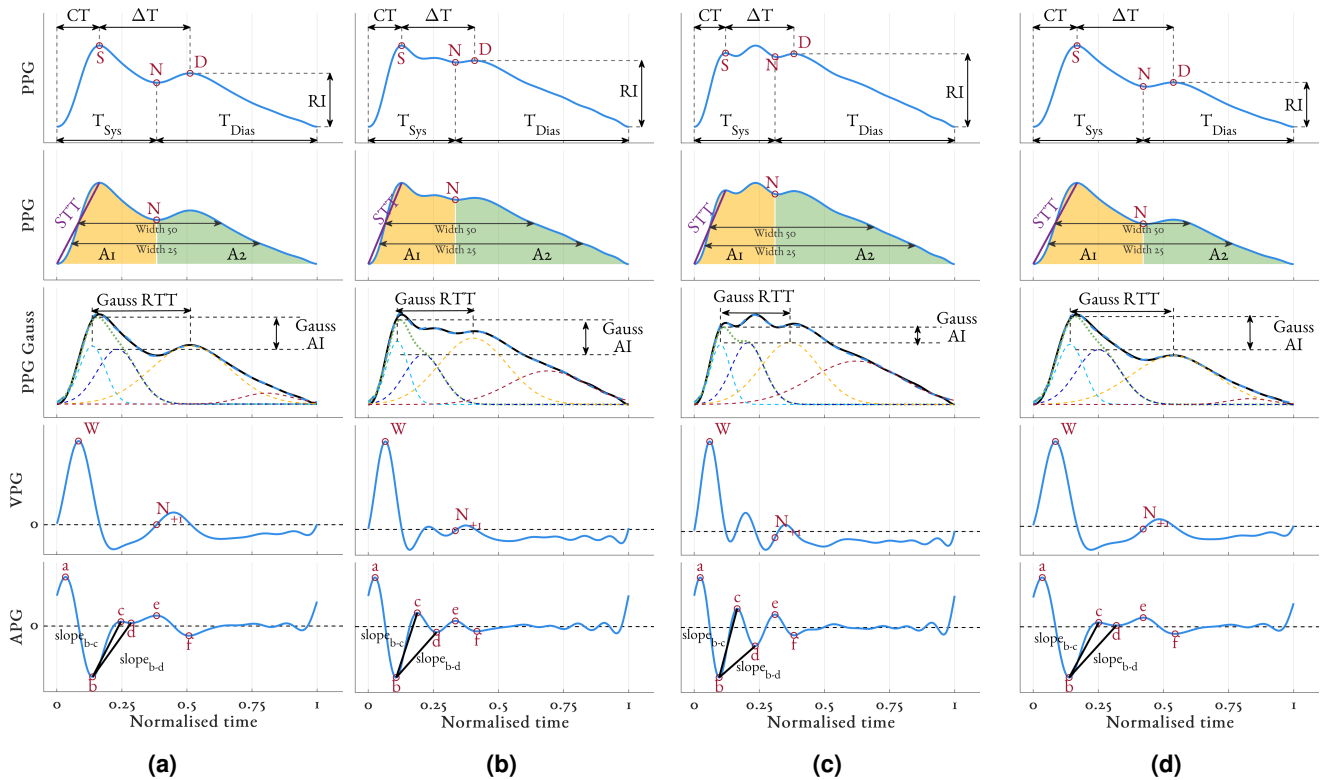

Figure 1. Overview of the fiducial points detected and some of the features extracted from the PPG waveform for one individual (Male, Age: 24, BMI: 25.1) during the four stages of the study protocol: (a) rest, (b) dose increase, (c) maximum infusion, and (d) washout. Examples of the following features are provided: Crest time (CT), ΔT, reflection index (RI), width$_{25}$, width$_{50}$, slope transit time (STT), A1, A2, Gaussian estimation of the transit time of the reflected wave (Gauss RTT) and augmentation index (Gauss AI), slope$_{b-c}$ and slope$_{b-d}$. Acronyms: S - Systolic peak, N - Dicrotic notch, D - Diastolic peak, a-e - waves of the APG.

Table 2. Summary of features extracted from the PPG.

Category	Feature notation and publication	Description or formula
PPG morphology	Notch amplitude (N_{amp})	Amplitude of the dicrotic notch (see figure 1)
	Reflective index (RI)[36]	Amplitude of the diastolic peak (see figure 1)
	ΔT[37]	Time from systolic peak to diastolic peak (see figure 1)
	Crest Time (CT)[13]	Time from onset to systolic peak (see figure 1)
	T_{Sys}[38]	Time in systolic phase (see figure 1)
	T_{Dia}[38]	Time in diastolic phase (see figure 1)
	T_{Ratio}[38]	T_{Sys}/T_{Dia}
	Slope transit time (STT)[39]	Slope of straight line from onset to peak (see figure 1)
	Stress-Induced Vascular Response Index (sVRI)[40]	μ_{Dias}/μ_{Sys}
	A1[41]	Area under PPG in systolic phase (see figure 1)
	A2[41]	Area under PPG in diastolic phase (see figure 1)
	Inflection point area (IPA)[41]	A2/A1
	$Width_{25}$[37]	Width of the PPG at 25 % of its amplitude (see figure 1)
	$Width_{50}$[37]	Width of the PPG at 50 % of its amplitude (see figure 1)
	Pressure index (PI)[42]	$\frac{t(N)-t(S)}{t(N)-t(W)} \times h$
	Normalised Harmonic Area (NHA)[41]	$\sum_{n=2}^{N} FFT^2(f_n)/\sum_{n=1}^{N} FFT^2(f_n)$
	Inflection and Harmonic area ratio (IHAR)[41]	(1-NHA)/ IPA
	Skewness[21]	–
	Kurtosis[21]	–
VPG morphology	Sys_μ[16]	Mean of VPG in the systolic phase
	Sys_σ[16]	Variance of VPG in the systolic phase
	Dia_μ[16]	Mean of VPG in the diastolic phase
	Dia_σ[16]	Variance of VPG in the diastolic phase
APG morphology	$\frac{b}{a}, \frac{c}{a}, \frac{d}{a}$ & $\frac{e}{a}$[43]	
	Ageing index (AGI)[43]	$\frac{b-c-d-e}{a}$
	$slope_{b\text{-}c}$[38]	Slope of a straight line between b and c, normalised by a (see figure 1)
	$slope_{b\text{-}d}$[38]	Slope of a straight line between b and d, normalised by a (see figure 1)
	PPG AI[44]	PPG augmentation index d_{-2}/b_{-2}
Gaussian decomposition	$A_{g1-4}, \sigma_{g1-4}, \mu_{g1-4}$	Amplitude, variance and mean of the four decomposed Gaussians
	Gaussian augmentation index (Gauss AI)[10]	$\max(g_s) - A_{g3}$ (see figure 1)
	Gaussian reflection index (Gauss RI)[10]	$\sum(g_s) - \sum(g_3)$
	Gaussian reflected wave transit time (Gauss RTT)[45]	$\mu_{g3} - \mu_{g1}$ (see figure 1)
	Gaussian augmentation index$_R$ (Gauss AI$_R$)[45] †	$\frac{A_{g1}-A_{g2}}{A_{g1}}$
	Gaussian reflection index$_R$ (Gauss RI$_R$)[45] †	A_{g3}/A_{g1}
	Gaussian approximation of left ventricular ejection time (Gauss LVET)[46]	See reference for definition
	Gauss$_{Sys/Dias}$ *	$\sum(g_s)/\sum(g_d)$
	Gauss$_{A4/A1}$ *	A_{g4}/A_{g1}
	Gauss$_{\sigma4/A1}$ *	σ_{g4}/A_{g1}
PCA	PPG PCA$_{1\text{-}3}$[47]	First 3 principal components of PPG beat
	VPG PCA$_{1\text{-}3}$[47]	First 3 principal components of VPG beat
	APG PCA$_{1\text{-}3}$[47]	First 3 principal components of APG beat

* Indicates features that, to the authors' knowledge, have not been previously implemented for BP estimation. † Note that there are two Gaussian indices referenced reflection index and augmentation index. We refer to the second set using the subscript R to reflect the authors: Rubins *et al.* μ_{Sys} and μ_{Dias} represents the mean of the PPG during systole and diastole respectively. h is the height of the participant. t is the time since the pulse onset. FFT is the fast Fourier transform. g_i refers to the ith Gaussian; $g_s = g_1 + g_2$ representing the systolic wave; $g_d = g_3 + g_4$ representing the diastolic wave. $\sum(g_i)$ is the area under the ith Gaussian.

2.2.1 Gaussian decomposition

Each PPG pulse (of unit amplitude and duration) was decomposed into the summation of four Gaussians[45]. The first two Gaussians represent the systolic component of the pulse, the second two Gaussians represent the diastolic component. This approach has the advantage of providing a representation of the PPG pulse without a reliance on fiducial point detection. Additionally, it allows for the modelling of reflected wave interactions which are thought to have a Gaussian profile[48]. For a PPG pulse ζ we computed the modelled pulse, ζ^{Gauss}, as:

$$\zeta^{\text{Gauss}}(t, \Theta) = \sum_{i=1}^{4} g_i(t, \theta_i) \tag{1}$$

where g_i represents the ith Gaussian component:

$$g_i(t, \theta_i) = A_{gi} \times e^{-\frac{(t - \mu_{gi})^2}{2\sigma_{gi}^2}} \tag{2}$$

t is the time duration for the PPG pulse, and θ_i is a vector, $[A_{gi}, \mu_{gi}, \sigma_{gi}]$, containing the respective amplitude, mean and variance of each Gaussian. $\Theta = [\theta_1, \theta_2, \theta_3, \theta_4]$ and thus Gaussian decomposition parameterises each PPG beat into 12 components. To determine the optimum value for Θ, $\hat{\Theta}$, we implemented a bounded Levenberg-Marquart optimisation algorithm to minimise the root mean squared error loss, L^{Gauss}, between ζ and ζ^{Gauss} given as:

$$L^{\text{Gauss}}(\Theta) = \sqrt{\frac{1}{N} \sum (\zeta - \zeta^{\text{Gauss}}(t, \Theta))^2} \quad , \qquad \hat{\Theta} = \underset{\Theta}{\text{argmin}}(L^{\text{Gauss}}(\Theta)) \tag{3}$$

The optimisation was bounded such that all parameters were positive and the amplitudes were all less than 1. L^{Gauss} is non-convex and therefore the optimised values were dependent on initial conditions. For the first beat, the initial conditions were: $\theta_1 = [1/2, 0.2, 0.01]$, $\theta_2 = [1/3, 0.4, 0.01]$, $\theta_3 = [1/4, 0.6, 0.01]$, and $\theta_4 = [1/6, 0.8, 0.01]$. To encourage continuity of parameters from beat-to-beat, we used the optimised parameters for the previous beat as initial seeds for the optimisation of the current beat. Following the work of[49], we set the SQI_{PPG} of each beat to 0 if the value of $L^{\text{Gauss}}(\hat{\Theta})$ for that beat was greater than 0.03.

Table 2 provides a summary of the Gaussian decomposition features used. Together with the values of $\hat{\Theta}$, we implemented various features derived from the Gaussian decomposition that have been previously proposed as indicators of arterial stiffness in the literature[45, 46]. Additionally, through observations of Gaussian decomposition in our dataset, we propose three new features: the ratio of the systolic component to the diastolic component ($\text{Gauss}_{Sys/Dias}$); the amplitude of the fourth Gaussian scaled by the amplitude of the first Gaussian ($\text{Gauss}_{A4/A1}$); and the variance of the fourth Gaussian scaled by the amplitude of the first Gaussian ($\text{Gauss}_{\sigma4/A1}$) (scaling by variance rather than amplitude of the first Gaussian gave a less informative parameter). Figure 1 shows an example of Gaussian decomposition for a typical participant during the four main stages of the study protocol: rest, dose increase, maximum infusion, and washout. Figure 1 also presents examples of feature extraction for Gaussian estimation of the transit time of the reflected wave (Gauss RTT) and augmentation index (Gauss AI).

2.2.2 Principal components

Principal component analysis (PCA)[50] maps high-dimensional data to a lower dimension along orthogonal principal components. These principal components account for the majority of the variation in the original data and therefore highlight regions of significant change in the PPG, VPG and APG signals. We computed PCA features using the following steps:

1. Resample all good-quality beats (defined as an $\text{SQI}_{\text{PPG}} > 0.8$) from the PPG, VPG and APG signals to be 100 samples in length using cubic spline interpolation.

2. Pool all resampled PPG, VPG and APG beats from all participants to form 3 matrices: Ψ_{PPG}, Ψ_{VPG} and Ψ_{APG} respectively.

3. Mean normalise each Ψ matrix.

4. Perform PCA on each Ψ independently by computing the eigenvectors of the corresponding covariance matrix and extract the first 3 principal components that correspond to the largest eigenvalues.

We extracted 3 principal components as this was found empirically to explain more than 85% of the variation in the Ψ_{PPG}, Ψ_{VPG} and Ψ_{APG} datasets. A visualisation of the computed PCA eigenvectors is provided in Supplementary Information SI: 1.

2.3 Features from the ECG

To preprocess the ECG waveform, we performed the following steps. To suppress the impact of baseline wander, the ECG was filtered using an 8th-order Butterworth IIR high-pass filter with a cut-off frequency of 0.5 Hz. To suppress power-line interference, a 2nd-order IIR notch filter with centre frequency at 50 Hz (the frequency of mains power in the UK) was used. We detected the QRS complex following the work of Pan and Tompkins[51] and assessed the quality of the ECG (SQI_{ECG}) following the work of Li *et al.*[52].

The features we extracted from the ECG are summarised in table 3. Features relating to complexity and entropy of the ECG have been previously proposed to track changes in BP[26,53]. These features quantify the level of regularity and unpredictability of fluctuations over a time series. Generally, a higher-level complexity indicates a more irregular dynamic system. A lower-level complexity indicates the presence of central trends or cyclical patterns. Changes in entropy of the ECG time series have been shown to track changes in heart rate variability (HRV) caused by myocardial ischaemia[54] and also denote periods of cardiac arrhythmia[55]. We provide full details of the algorithms used in Supplementary Information SI: 2.

Table 3. Summary of features extracted from the ECG.

Feature notation and publication	Description or formula
Hjorth mobility[26]	Estimate of the signal's mean frequency (Equation SI 1)
Hjorth complexity[26]	Estimate of the signal's bandwidth (Equation SI 2)
Fractal dimension[26]	Computed using Higuchi's algorithm[56] with $k_{max} = 17$
Shannon entropy (SE)[26]	Uncertainty of information content based on probability distribution (Equation SI 3)
Approximate entropy (approxEnt) *	Quantifies regularities of signal
Sample entropy (sampEnt) *	approxEnt computed without self comparisons
Multi-level sample entropy (MSE) *	Course approximations of sample entropy, computed at scales 2, 4, 6, 8

* Indicates features that, to the authors' knowledge, have not been previously implemented for BP estimation. Full details of the algorithms and equations used for computing each feature can be found in Supplementary Information SI: 2.

2.4 Pulse Arrival Time

PAT has been shown previously to provide a beat-by-beat estimate of changes in arterial stiffness and therefore may be a good surrogate for BP[1]. PAT and its corresponding SQI_{PAT} was computed in the same manner as we have previously reported[1]. We used PAT in a baseline model to compare the performance of PPG and ECG features for BP estimation.

2.5 Computing the reference BP values

Measurements of SBP, MAP, and DBP using sphygmomanometer cuffs have known limitations depending on posture, cuff-inflation hypertension and cuff size[57,58]. The sphygmomanometer cuff used in our study was programmed to inflate every minute. However, errors in cuff inflation prevented the Philips monitor from registering an accurate estimate, resulting in a missed data-point in the recorded BP time series. Therefore, data from the cuff was both noisy and sparse. In order to reduce the impact of these sources of error, we processed the cuff data using a cubic smoothing splines[59] algorithm. This allowed for both filtering and interpolation of the noisy blood pressure readings to a new sampling frequency f_{BP} set as once per minute. Cubic smoothing spline defines a function, $\hat{f(t)}$, that minimises a loss L^{BP}:

$$L^{BP} = p \sum (y_i - f(\hat{t_i}))^2 + \int f''\hat{(t)}^2 dt \tag{4}$$

where y_i are the observed BP values from the cuff, occurring at time t_i, $f''\hat{(t)}$ is the second derivative of $\hat{f(t)}$. p is a constant that defines the relative weight placed on minimising the residual sum of squares against the 'roughness' (the accumulated second derivative) of the regressed signal $\hat{f}$. As all participants were studied under the same protocol, we implemented p values for SBP, MAP and DBP (p_{SBP}, p_{MAP} and p_{DBP} respectively) that were common for all participants in the dataset. More details of the methods for computing the reference BP values can be found in Supplementary Information SI: 3.

2.6 Estimating changes in BP

We processed PPG features and PAT similarly to the methods proposed in[1]. This included: outlier detection to remove statistically significant deviations in values and a Kalman filter to reduce the effect of transient artefacts caused by noise. We then averaged the feature values within a window, w, of length 40s centred around each reference BP measurement (20s to the left, 20s to the right). Only beats of good quality, given by $SQI_{PPG} > 0.8$ and $SQI_{PAT} > 0.8$ respectively, were included in the window and if more than half of the window was deemed to be of bad quality, then the feature was not recorded for that window. We computed ECG features within the same window, w. If less than half of the window was deemed to be of good quality ($SQI_{ECG} > 0.8$), then feature values were not recorded for that window. We handled missing data, caused by poo signal quality, by nearest neighbour imputation for each participant.

A schematic outlining our proposed steps for estimating changes in ΔBP is shown in figure 2. In this work, we adopted a *hybrid calibration*[60] approach in order to estimate ΔBP using one of two ML regression models (LASSO+OLS or RF, defined in the sections below) as a function of an input feature set, X. We implemented four different feature sets based on the different signals being analysed in this study. For each of the following groups we restricted X such that it includes features only from these sources: $X \in \{PPG, ECG, PPG+ECG, PAT\}$.
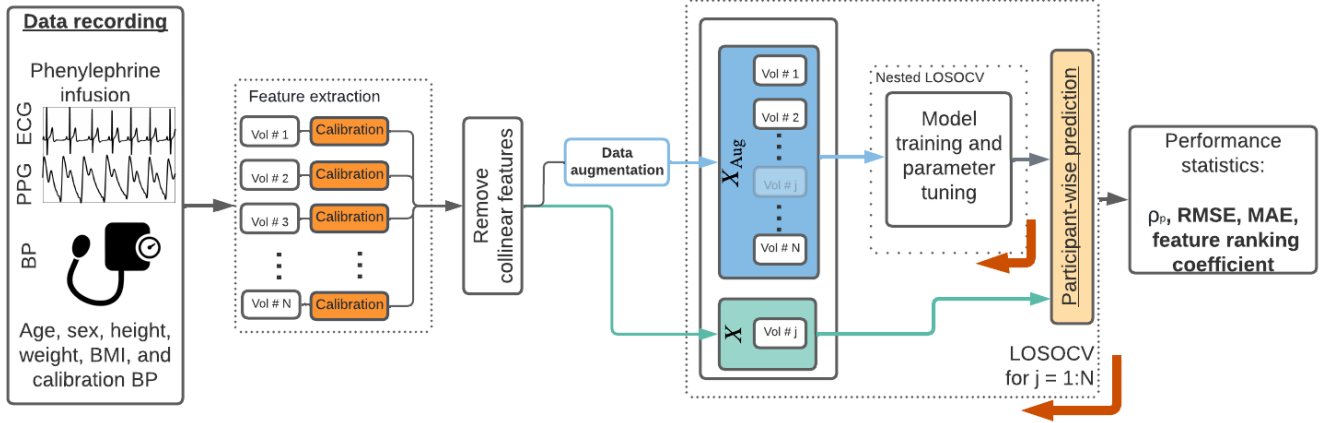
Figure 2. Schematic of the ΔBP estimation pipeline for each of the proposed models. We extracted features from the PPG and ECG and averaged their values within a window of size 40s centred on times of cuff inflations. We then implemented a *hybrid calibration* approach such that the proposed models estimate ΔBP from a baseline calibration value determined during the rest period. Data augmentation was implemented to increase the training and validation set size by interpolating between cuff inflations. Models were trained and evaluated in a nested leave-one-subject-out cross-validation (LOSOCV) framework shown here by the iterator j which indicates the test participant for that iteration. Participant j was then removed from the training/validation set (X_{Aug}) for that iteration.

We refer to these models and feature set combinations as LASSO + OLS$_{\text{PPG}}$ for a LASSO + OLS model with a PPG feature set, RF$_{\text{PPG + ECG}}$ for a RF model with PPG + ECG feature set, and so on.

We calibrated all feature and BP values to the individual participant using data recorded during the rest period of the study (first 5 minutes of the recording). We then removed all collinear features and implemented data augmentation to increase the size of the training set, X_{Aug}. We used nested leave-one-subject-out cross-validation (LOSOCV) to train, tune and evaluate the models. For each fold of the LOSOCV, one participant in turn was set as the test participant. Data from X_{Aug} set for all participants apart from the test participant were used to train and tune the models. Nested LOSOCV was used for model validation to optimise model hyperparameters (λ for OLS+LASSO and *mtry* for RF) with the aim of minimising RMSE. Data from X for the test participant was then used to evaluate the performance of the model. We used the following metrics to evaluate model performance: ρ_p, RMSE, and mean absolute error (MAE). Differences in performance metrics across all folds from the models were evaluated for statistical significance by a two-tailed Wilcoxon signed-rank test. We adjusted the *p*-values for multiple comparisons using the Benjamini-Hochberg method[61]. This technique aims to control the number of type I errors (incorrectly rejecting the null hypothesis) by inflating the lowest *p* values (see[61] for more details).

2.6.1 Calibration

We adopted a *hybrid calibration*[60] approach to personalise the estimation models for each participant in our dataset. In this work, we defined the baseline calibration value of BP (BP$^{\text{calib}}$) and waveform features (f^{calib}) as their respective mean values during the 5-minute resting period at the start of the study. In the specific example of one participant for whom no BP readings were taken during the rest period, we used the mean values in a one-minute window centred on the first cuff inflation as the calibration readings. For each participant, BP values were subtracted from their calibration value. Relative changes in feature values from their calibration values were used[62]:

$$\Delta \text{BP} = \text{BP} - \text{BP}^{\text{calib}}, \qquad \Delta f = \frac{f - f^{\text{calib}}}{f^{\text{calib}}}$$

For all feature sets, X, we added participant age, sex, height, weight, BMI and BP$^{\text{calib}}$ as static categorical features.

2.6.2 Baseline reference

For reference performance metrics, we implemented a simple baseline reference that assumed no BP changes for each participant from their baseline calibration value BP$^{\text{calib}}$ (i.e. ΔBP = 0). We refer to this as *Baseline reference*.

2.6.3 Removing collinear features

Collinearity occurs when there is intercorrelation between multiple features[63,64], thus violating the independent identically distributed (i.i.d.) assumption that is common in regression models. Additionally, the presence of collinearity inflates the variance of the regression parameters and makes it difficult to assess the importance of features. Collinearity in a feature set, X, can be highlighted by the condition number κ representing the ratio of the largest singular value of X to the smallest singular value. It can be computed as:

$$\kappa(X) = \left\| X^+ \right\| \cdot \left\| X \right\| \tag{5}$$

where $\|\cdot\|$ is the 2-norm of a matrix and X^+ is the pseudo inverse of the matrix X. Typically, a condition number greater than 30 is thought to indicate the presence of strong multi-collinearity in the dataset[64].

It is likely that collinearity exists in the feature sets presented in this paper as there are multiple features describing similar characteristics, for example entropies of the ECG or time durations of the PPG. In order to remove the effect of collinearity and to allow for parsimonious models, we removed collinear features by investigating the variance inflation factor (VIF) defined as:

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \tag{6}$$

where R_i^2 is the unadjusted coefficient of determination for regressing the ith feature on the remaining ones. Removing collinear features is an iterative process where on each iteration, the feature with the largest corresponding VIF_i is removed from the feature set until no features had a VIF greater than 10 (corresponding to $R_i = 0.9$)[64].

2.6.4 Data augmentation

We implemented data augmentation in order to increase the feature set size by incorporating information from feature values between the reference BP cuff inflations. As shown in figure 2, we performed model training and validation on the augmented feature set, hereafter referred to as X_{Aug}, and performance metrics were reported using the original dataset X. We constructed X_{Aug} by interpolating between the reference BP values for each participant using the cubic smoothing splines outlined above at a new frequency, $f_{\text{BP}} = 1/15$ Hz (four measurements per minute, as opposed to once a minute) with a smaller window size, $w = 15$s, to prevent overlapping windows violating the i.i.d assumption.

2.6.5 Regression models

The models we implemented to estimate ΔBP are outlined below. Different models were built for estimating ΔSBP, ΔMAP and ΔDBP.

LASSO + OLS   To explore the linear relationship between each of the feature sets X and ΔBP, ordinary least squares (OLS) linear regression was implemented. In order to prevent over-fitting and to improve model interpretability, we employed the Least Absolute Shrinkage and Selection Operator (LASSO) method to remove redundant features prior to linear regression. We refer to this model as LASSO+OLS. LASSO imposes the L1-norm penalty to the residual sum of squares using non-negative values of shrinkage parameter λ. LASSO allows removal of features by shrinking some feature coefficients, β, towards zero:

$$\beta = \underset{\beta}{\text{argmin}}\left(\sum_{i=1}^{N}(Y_i - X_i\beta)^2 + \lambda\sum_{j=1}^{M}|\beta_j|\right) \tag{7}$$

where Y is the target vector (ΔBP values) of length N, and M is the number of features. $|\beta_j|$ can therefore provide an indication of the relative importance of each feature. We optimised the λ hyper-parameter by a nested LOSOCV loop. For each loop of the LOSOCV, LASSO feature selection was implemented and features with non-zero coefficients were used by OLS to compute ΔBP estimates.

Random forest   To explore potentially non-linear relationships between X and ΔBP, we additionally built a Random Forest (RF) regression model. RF regression models utilise majority voting across multiple decision trees, each trained with a split criterion based on summed squared error (SSE)[65]. Each decision tree in a RF model was trained on a bootstrap of features. This approach reduces model variance whilst maintaining a low bias. As RF models select features upon training, we trained the model using all available features (i.e. without the need for LASSO). Typically, RF models are not very sensitive to choices in the number of trees (N_{trees}), provided it is sufficiently high[65]. Therefore, we set the number of trees to 300. We optimised the number of features randomly selected for each node (labelled *mtry*) by a nested LOSOCV loop. The relative importance of each feature was assessed as the decrease in SSE averaged across all nodes and all trees that split based on that feature (see Genuer *et al.*[65] for more details).

2.6.6 Feature ranking coefficient

A key objective of this work was to highlight features that have strong predictive power for estimating ΔBP. For each loop of the cross-validation (CV), features may have different relative feature importance values (provided by either LASSO or RF feature selection algorithms). To assess the variability of the relative importance, we computed a ranking coefficient. The ranks of all features were determined at each fold and normalised by the total number of features (1 being the highest rank, 0 being the lowest). For each feature, the distributions of the ranking coefficients across all folds were analysed.

3 Results

3.1 Clinical study

Thirty volunteers were recruited for our clinical study. We discarded the data from four participants from the analysis. For three of these participants, the reference ECG waveform did not include any periods of high-quality data as a result of errors in the connection of the

ECG electrodes. For one participant there were errors recording the BP cuff data. Therefore, 26 participants made up our dataset. The demographics of the participants in the study whose data was used for analysis are shown in table 4. All participants were healthy with a median BMI of 22.5 kg/m^2 and no history of cardiovascular disease. The median age of participants was 28 years and there was an even split of sexes (53.8% female). On average, we achieved an increase of 20 mmHg in SBP, with a maximum increase of 40 mmHg in a subset of participants.

Table 4. Demographics of the population in the clinical study.

Descriptor	Value
Total number of participants	26
Male	12
Female	14
Average length of session (mins)[1]	28.0 (0.1)
Age (years)[1]	28 (9.0)
Height (cm)[1]	170.0 (18.0)
Weight (kg)[1]	69.5 (23.0)
Body Mass Index (kg/m^2)[1]	22.5 (5.2)
Reference maximum Δ BP per participant[1]	
Δ SBP (mmHg)	20.0 (8.0)
Δ MAP (mmHg)	17.0 (10.0)
Δ DBP (mmHg)	15.5 (9.0)
Reference Δ BP values[2]	
Δ SBP (mmHg)	6.5 (10.4)
Δ MAP (mmHg)	5.3 (8.8)
Δ DBP (mmHg)	4.7 (8.2)

[1] Participant-wise median (interquartile range (IQR)), [2] mean (standard deviation) across dataset.

3.2 Removing collinear features

Figure 3 (a) shows the correlation matrix of the total feature set (PPG + ECG + demographics), including 77 features. There were a large number of features (58.5%) which were significantly correlations ($|\rho_p| > 0.8$, $p < 0.05$) with at least one other feature. This indicates a high level of collinearity, highlighted by a condition number κ of 315. Figure 3 (b) shows the correlation matrix after removing all collinear features. The remaining dataset contained 45 features with a condition number κ of 11, suggesting independence of features and encouraging parsimonious models. For completeness, Supplementary Information table SI 3 provides a list of the remaining features and the subset of the total feature set with which they have a strong correlation. Additionally, Supplementary Information table SI 3 provides the correlation with ΔSBP for each feature across the whole cohort and on a participant-wise basis.

3.3 Comparing model performance

Table 5 shows the median and interquartile range (IQR) performance statistics for all models computed over all 26 folds of the LOSOCV for ΔSBP. The results for ΔMAP and ΔDBP are provided in the Supplementary Information table SI 1 and table SI 2 respectively. We note large RMSE and MAE values for the *baseline reference* indicating that participants experienced a significant change in their SBP values in response to the weight-based dosing of phenylephrine.

All models outperformed results obtained with the *baseline reference*, and all apart from LASSO + OLS$_{ECG}$ reported statistically significant p values indicating that consistent improvements in performance statistics were observed ($p < 0.05$ for all). For the PPG, ECG and PPG+ECG feature sets, the RF model consistently achieved stronger performance metrics than LASSO + OLS. Statistically significant improvements were recorded only in RMSE and MAE ($p < 0.05$ for all). The Wilcoxon signed rank test failed to reject the null hypothesis of equal median ρ_p between LASSO+OLS and RF. The PPG feature set significantly outperformed the ECG feature set for all performance metrics and regression models. The absolute difference in the median ρ_p, RMSE, and MAE between RF$_{PPG}$ and RF$_{ECG}$ was 0.19 ($p = 0.00007$), 1.04 mmHg ($p = 0.005$) and 0.53 mmHg ($p = 0.004$) respectively. RF$_{PPG}$ and RF$_{PPG + ECG}$ reported similar performance statistics with non-significant p values, indicating that adding ECG features to a feature set of PPG features offers little or no performance gain. For PAT, LASSO + OLS significantly outperformed RF across all performance metrics ($p < 0.05$ for all). LASSO + OLS$_{PAT}$ achieved similar performance metrics to the RF$_{PPG + ECG}$. LASSO + OLS$_{PAT}$ consistently resulted in the smallest IQR for all performance metrics.

Figure 4 shows the (a) correlation and (b) Bland-Altman plots for ΔSBP estimation using the RF$_{PPG + ECG}$ model. Individual participants are colour and marker-coded. The ρ_p value of the overall estimation was 0.64. The median participant-wise correlation coefficient was 0.86 with a range of 0.34 to 0.95. Figure 4 (a) shows the histograms of the reference and estimated values. The reference ΔSBP values ranged from -16.4 to 53.8 mmHg, but the estimated ΔSBP values had a much tighter range of -3.37 to 22.2 mmHg. The bias of the overall error was 0.30 mmHg with a standard deviation of 8.05 mmHg (see figure 4 (b)). We note also an additional bias where large values of ΔSBP were
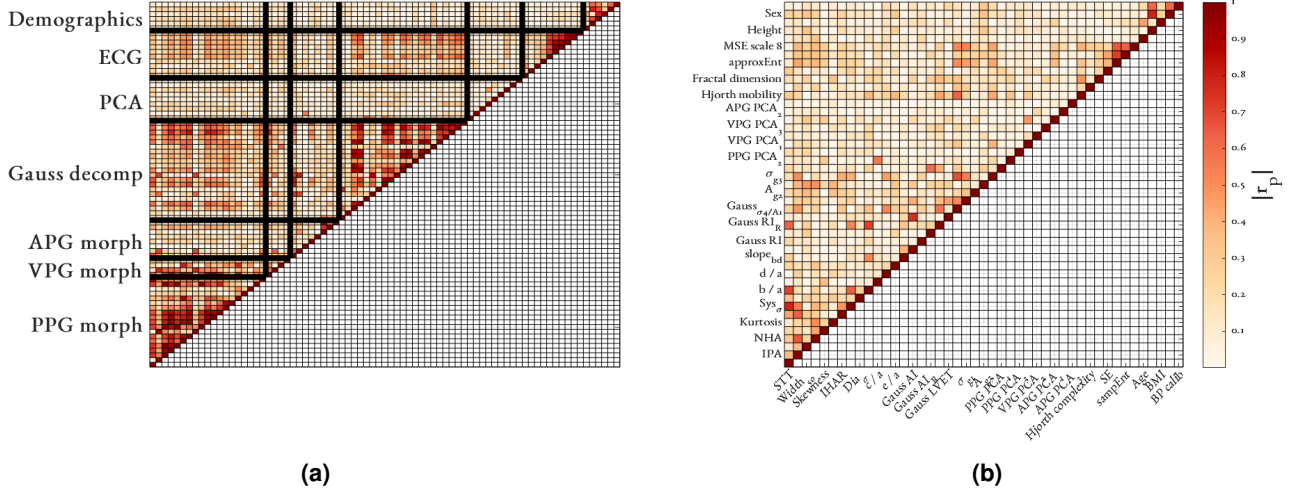
Figure 3. Results of removing collinear features in PPG + ECG feature set. Pairwise linear absolute Pearson's $|\rho_p|$ correlation matrix. (a) pre and (b) post removal of collinear features. Blank spaces represent non-significant ($p > 0.05$) or weak ($|\rho_p| < 0.2$) correlations. For clarity of labelling on the post-collinear feature set (b), every other label is presented on the y-axis with the remaining labels on the x-axis.

underestimated and small values were overestimated. At peak infusion, the median (IQR) value of ΔSBP across the cohort was 20 (8) mmHg (see table 4). We found large errors in the data for the four participants whose ΔSBP at peak infusion exceeded 30 mmHg.

Supplementary Information figures SI 3-4 shows the individual reference ΔBP and estimated ΔBP values using the $\text{RF}_{\text{PPG + ECG}}$ model across all participants in the study for SBP, MAP and DBP respectively.

Table 5. Performance statistics of ΔSBP estimation using the models proposed. Results are given as median (IQR) computed across all folds of the LOSOCV. Entries in bold indicate the best performance for that metric.

Model name	ρ_p		RMSE [*]		MAE [*]	
Baseline reference	-	(-)	10.46	(6.18)	7.40	(4.87)
LASSO+OLS$_{\text{PPG}}$	0.85	(0.44)	6.19	(4.05)	5.41	(2.41)
RF$_{\text{PPG}}$	**0.88**	**(0.26)**	5.75	(4.73)	**4.75**	**(3.66)**
LASSO+OLS$_{\text{ECG}}$	0.64	(0.70)	7.48	(6.37)	6.54	(4.63)
RF$_{\text{ECG}}$	0.69	(0.45)	6.79	(5.66)	5.28	(4.57)
LASSO+OLS$_{\text{PPG+ECG}}$	0.82	(0.40)	7.99	(5.74)	6.30	(4.62)
RF$_{\text{PPG+ECG}}$	0.86	(0.23)	**5.66**	**(4.76)**	4.86	(4.29)
LASSO+OLS$_{\text{PAT}}$	0.83	(0.10)	6.09	(4.25)	4.91	(3.33)
RF$_{\text{PAT}}$	0.80	(0.15)	6.39	(4.23)	5.21	(3.69)

[*] results given in units of mmHg

3.4 Feature importance

To mitigate the variations in the training data between folds of the LOSOCV, overall feature importance was assessed through using a ranking coefficient. Figure 5 shows the median (across folds) RF and LASSO ranking coefficients for the PPG+ECG feature set, with only the top 15 features shown (in figure 5, all features are shown in Supplementary Information figure SI 2). The top five features as determined by the median RF ranking coefficient were: kurtosis, Gauss LVET, e/a, c/a, and BMI. The top five features as determined by the median LASSO ranking coefficients were: Gauss$_{\sigma4/A1}$, kurtosis, Gauss LVET, age, VPG PCA$_3$. Kurtosis had the highest feature importance in 22 out of 26 folds for RF and Gauss$_{\sigma4/A1}$ had the highest feature importance in 19 out of 26 folds for LASSO. In general, the ranking of features by LASSO followed the order of feature correlation values shown in Supplementary Information table SI 3. LASSO put a large emphasis on all demographic features, whereas RF put a comparatively low emphasis on height and sex. For RF, ECG features had very low importance with the strongest performing feature, sampEnt, having a median ranking coefficient of 0.35. LASSO placed slightly larger importance on ECG features, with Hjorth mobility having a median ranking coefficient of 0.60.

Figure 6 shows the relationship between ΔSBP and the top 9 ranking, non-demographic, features from the $\text{RF}_{\text{PPG + ECG}}$ model. Different
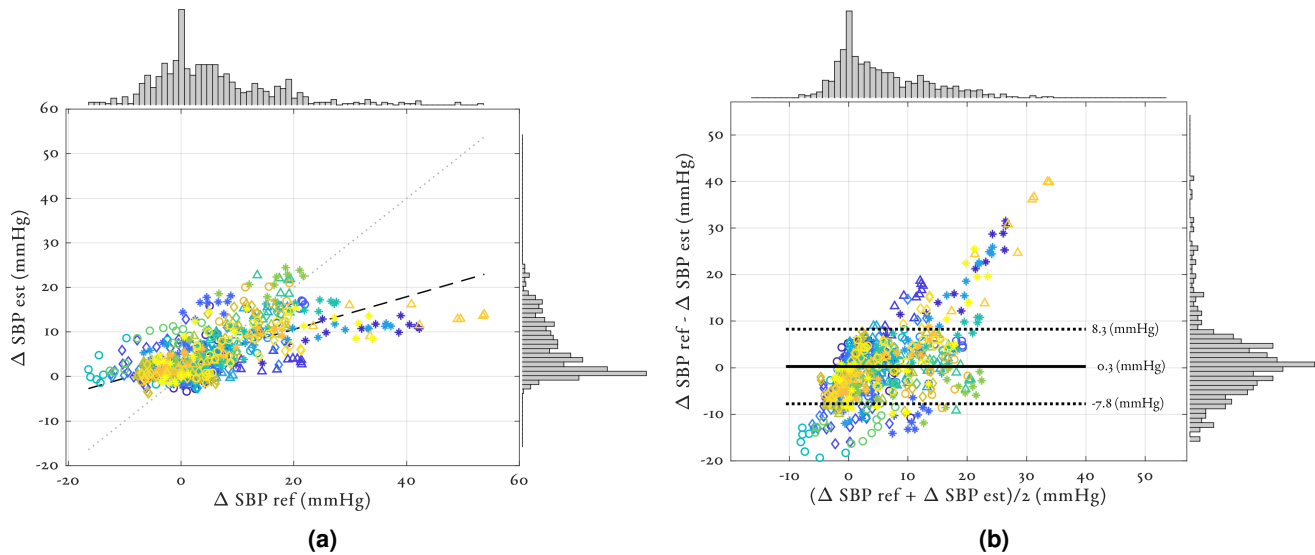
Figure 4. Agreement between the reference ΔSBP values from the sphygmomanometer cuff and the estimated ΔSBP using RF$_{PPG + ECG}$. Individual participants are colour and marker-coded. (a) The correlation analysis, the overall correlation was 0.64, the median participant-wise correlation coefficient was 0.86 with a range of 0.34 to 0.95. Black striped line shows regression line. Grey dotted line shows line of unity. (b) The Bland-Altman analysis, the bias of the overall error was 0.3 mmHg with a standard deviation of 8.05 mmHg

participants are coded by colour and marker to highlight clusters of features suggesting individual specific feature changes. Both Pearson's ρ_p and Spearman's ρ_s correlation coefficient are provided. Spearman's correlation indicates monotonic (but not necessarily linear) relationships and so may provide further insight into the RF models (see for example the scatter of $\sigma g1$).

4 Discussion

This work describes the methods for the non-invasive estimation of ΔBP in healthy participants using morphological features from the PPG and ECG waveforms. Changes in BP were induced by the infusion of phenylephrine using a weight-based dosing protocol, instead of being BP-target driven (although BP was constantly under review by the clinician to ensure safety of the participants). One of the key advantages of this study was that, in a relatively short period of time, and while remaining supine and still, the participants experienced a wide range of BP values. This helped to validate algorithms for non-invasive, cuffless, estimation of BP over a clinically useful range of ΔSBP (-10 to 30 mmHg).

The PPG and the ECG waveforms offer great potential for non-invasive monitoring due to their ubiquity and ease of acquisition. The PPG in particular can be acquired using wearable devices such as smartwatches or using video plethysmography[12]. BP estimation using the PPG has been studied in a number of papers[12], however the best features and models have remained unclear. A single-lead ECG may be recorded by three electrodes or via capacitive coupling[66]. The relationship between changes in BP and the ECG is governed by mechano-electric coupling (MEC). However, this connection is considered less robust to that of the relationship between the PPG waveform and BP. As a result, estimating ΔBP from the ECG waveform has been explored in less detail in the literature.

4.1 Observed changes in the PPG waveform morphology

Figure 1 shows the changes in PPG beat seen for a typical participant during the four main stages of our study (a) rest, (b) dose increase, (c) max infusion and (d) washout. This includes a loss of a clear dicrotic notch (which is produced by the closing of the aortic valve and therefore marks the end of systole and the beginning of diastole) and a rising middle peak (often referred to as a tidal wave[67] and thought to be caused by reflected waves at the renal arteries[8]) that can overshadow the initial peak seen at rest. These changes have been reported previously[16,37,48] and are thought to be due to increasing amplitude and speed of reflected waves due to arterial stiffening. Additionally, they have also been shown to occur for age related arterial stiffening[67]. Note that in figure 1 (c) at maximum infusion, the tidal wave has a peak greater than that of the original systolic peak (see figure 1 (b)). As the systolic peak is almost universally defined as the maximum of the PPG beat[35], this may lead to the tidal wave being incorrectly classified as the systolic peak. In which case, features such as CT, ΔT, and STT, which rely on accurate systolic peak detection would be significantly affected. To account for this, we detected the systolic peak as the first turning point of the PPG pulse above its midpoint, as can be seen in figure 1 (c). We found that this definition was acceptable for our short, single perturbation study with minimal motion artefacts.
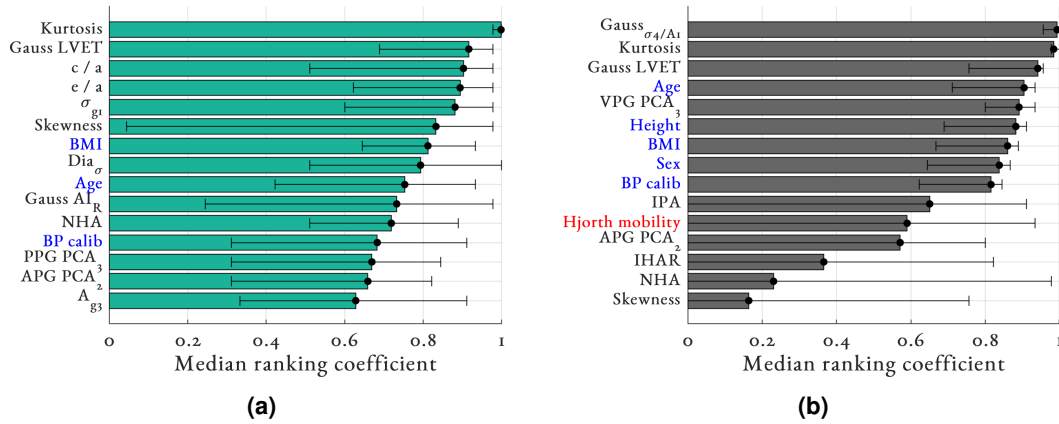
Figure 5. Median ranking coefficient for (a) random forest and (b) LASSO feature importance both with PPG + ECG feature set. Features are ordered by their respective median ranking coefficient. Error bars denote the range. Demographic features are highlighted in blue and ECG features in red for clarity. For brevity, only the top 15 features are provided here with the full list provided in Supplementary Information figure SI 2.

4.2 Model performance

Typically, machine learning algorithms require a large amount of data in order to make robust estimations. In this work, we were limited by the number of participants in the study (N=26) and by the number of data points per participant (typically once per minute for the 28 minutes of the recording per participant). In order to overcome these limitations, the models were trained, tuned and evaluated with a LOSOCV framework. Additionally, we implemented data augmentation to increase the number of data points in our training set by interpolating between the cuff measurements using cubic smoothing splines with an additional 3 augmented measurements every minute. Across all models, this data augmentation statistically significantly improved the performance statistics at the $p < 0.001$ level, computed using a Wilcoxon signed rank test. Figure 3 (a) splits all features into groups (PPG, VPG, APG, Gaussian decomposition, ECG and demographics) and shows strong intra-group correlations indicating multi-collinearity in the dataset. We reduced feature collinearity by removing features with a VIF > 10. This increased the confidence of the model parameters, allowed for appropriate examination of feature importance and encouraged parsimonious models.

For the PPG, ECG, and PPG+ECG feature sets, there were statistically significant improvements in RMSE and MAE when using the RF model compared to LASSO+OLS model at the $p < 0.05$ level as computed by the corrected Wilcoxon signed rank test, whereas a non-significant difference was found in ρ_p. This indicates that LASSO+OLS was able to appropriately detect directional changes in ΔSBP, although it was unable to determine the magnitudes of the changes. This points to a calibration issue whereby the LASSO+OLS model was unable to approximate the feature gradients for each participant. RF models may have been better at calibrating to ΔBP as these models allow for interactions between features and demographics. RF and LASSO both put a large emphasis on demographics (see figure 5); however, in linear models these static features can only influence intercepts and not gradients. For $RF_{PPG+ECG}$, a large importance was placed on BMI, age, and BP^{calib}, all of which have a reported strong relationship with arterial stiffness and calcification within the arterial wall[38, 68]. Demographics are known to influence the PPG contour; in particular, c/a has been previously explored as a marker for age-related arterial stiffening[37]. Interestingly, the $RF_{PPG+ECG}$ model puts little to no emphasis on the sex of the participant (see Supplementary Information figure SI 2) despite there being evidence of a sex-related dependency on the PPG morphology[14]. We would recommend in future studies that the interactions between participant demographics may be accounted for in a linear model by implementing a linear mixed-effects model with random effects for parameters such as age and BMI. We were unable to explore this line of work here due to the limited number of participants and their relative homogeneity.

Despite interactions with demographics, a significant calibration issue still persisted. Figure 4 shows the results of ΔSBP estimation using the $RF_{PPG + ECG}$ model via correlation and Bland-Altman analysis. The $RF_{PPG + ECG}$ model estimated ΔSBP values to be in a much tighter range than was given by the reference and there were large errors found particularly at the high values. The largest errors found in all models occurred in the four individuals with the largest ΔSBP at peak infusion (ids: 002, 010, 023 and 026 in Supplementary Information figure SI 3). These four individuals experienced significantly larger values of ΔSBP at peak infusion than the rest of the cohort (ΔSBP > 33 mmHg, whereas the median value across the cohort at peak infusion was 20 mmHg). The precise clinical effect of the weight-based dosing of phenylephrine in an individual would depend on the balance between their sensitivity to the increase in afterload, the effect of bradycardia on cardiac filling (therefore contractility, the Frank-Starling law[69]), and the proportion of venous/arterial action of phenylephrine causing the increase in preload and afterload. As a result, variations in changes of BP across the cohort were expected. The median weight of the four outlier individuals (73 kg) was marginally larger than the cohort average (69.5 kg) and so on average a larger dose of phenylephrine would have been given. Other post-hoc assessments of demographics are not sufficient to distinguish these four individuals. Although in such a small cohort it is difficult to draw conclusions for these individuals, we suggest two possible explanations for the large errors found. Firstly, these individuals clearly experienced a significant change to their cardiovascular system in response to the dosing of phenylephrine and so the changes in the cardiovascular system may not be adequately represented by the features available. Secondly, the *hybrid calibration*
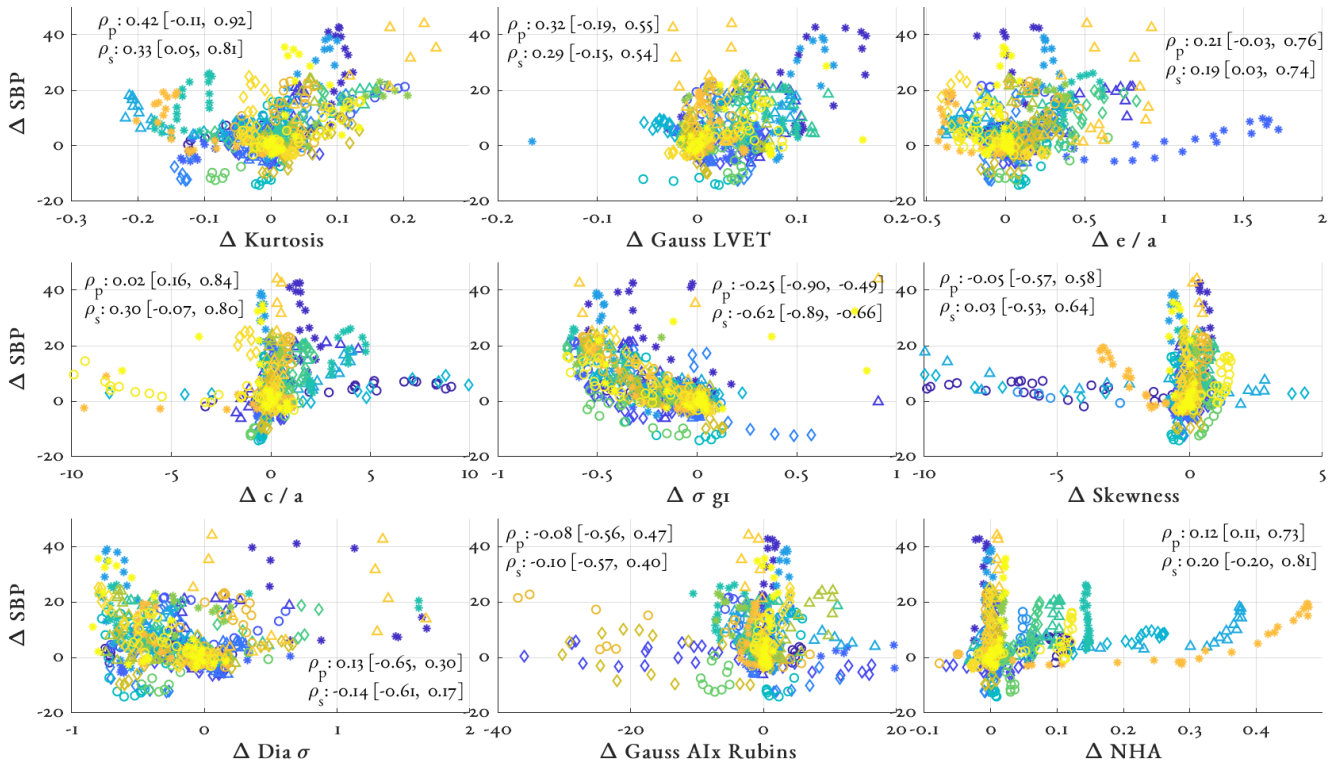
Figure 6. Relationship between top 9 ranking, non-demographic, features from the $RF_{PPG+ECG}$ model and ΔSBP. Individual participants are colour and marker coded. Inter Pearson's ρ_p and Spearman's ρ_s correlation coefficient are shown as median and upper and lower quartiles. Additionally the 1st and 3rd quartiles of the participant-wise correlation coefficients are shown. Note that feature values have been both subtracted and divided by their value at baseline.

strategy may be impacted by the small sample size (N=26). As a result, there will be difficulty in calibrating individuals that may be classed as outliers.

Likely driven by the linear relationship between PAT and BP[1], LASSO + OLS_{PAT} had stronger performance metrics than RF_{PAT}. PAT has been investigated as a surrogate measure of BP in this dataset previously[1] when we reported that *individual calibration*, as opposed to population-based *hybrid calibration* models, were needed for appropriate estimation of BP. In this work, there were insignificant differences found between $RF_{PPG+ECG}$ and LASSO + OLS_{PAT} suggesting that features from the PPG and ECG have the same calibration constraint as PAT. Estimating BP from the ECG and/or PPG may hold two significant advantages over PAT. Firstly, the devices do not need to be perfectly synchronous and recorded on the same internal clock as is required for computing PAT estimates and has been reported as a limitation in certain datasets[70]. It would be possible to design a system for which BP was estimated from both devices when they are available and only from one device when the other was disconnected for any reason. Secondly, BP estimation from the PPG and ECG features were not impacted by the pre-ejection period (PEP) which is a significant limitation to BP estimation using PAT[2].

The results in table 5 suggest that PPG features have a significantly stronger relationship to changes in BP than the complexity features we extracted from the ECG. Additionally, when adding the ECG to a PPG feature set, no performance improvements were observed. As mentioned in the introduction, the theoretical relationship between changes in BP and the ECG is governed by MEC and the poor performance suggests that external, non-cardiac, control mechanisms may have a significant impact on the BP-ECG relationship, affecting the latter's ability to estimate ΔBP. An additional explanation for the poor performance of the ECG feature set may be the choice of features used to explain the changes in ECG morphology; however to the authors' knowledge, there is no other work suggesting alternatives to ECG complexity features for BP estimation. For RF_{ECG}, improvements from a naive *baseline reference* assuming constant BP values were observed, but in general we suggest that ECG features on their own may not offer a viable solution to cuffless BP monitoring.

4.3 Feature importance

We explored a large and comprehensive pool of features from both the PPG and the ECG gathered from a wide range of previous work (see tables 2 and 3). Supplementary Information table SI 3 shows the features remaining for analysis after removing collinear features and the features from the original set with which they best correlate (defined as $|\rho_p| > 0.8$ $p < 0.05$). The overall correlations of these features to ΔSBP across the cohort was in general quite low with only one feature ($Gauss_{\sigma 4/A1}$) having $|\rho_{p\Delta SBP}| > 0.5$. There were a number of features that had significant participant-wise correlations (PWC) to ΔSBP, 21 features with a median absolute VWC > 0.5. The disparity between low correlations across the cohort and high correlations on a participant-wise basis underpins the important need for *individual calibration* due to

low intra-participant variability and high inter-participant variability.

BP is determined by cardiac output (CO) and total peripheral resistance (TPR), and changes in either of these may be represented by different features[71]. Phenylephrine causes a direct increase in TPR[72] via an increase in both arterial and venous vasoconstriction. Therefore, a large impact from the reflected waves, caused by impedance mismatches at points along the arterial tree (specifically the renal and iliac arteries[8]), was not unexpected. This was reflected in the observation that the majority of features that in this study have either a strong correlation or importance in estimating ΔSBP, characterise the impact of the reflected pressure waves (IPA, c/a, VPG PCA$_3$). On the other hand, phenylephrine causes a mixed response in CO with the relationship governed by preload dependency[72]. The majority of participants in the study experienced a decrease in CO (see Supplementary Information figure SI 6), driven largely by a decrease in heart rate. Therefore, it is not surprising that at least one of the most important features (σ_{g1}) represents changes in the upslope of the PPG which is driven by changes in CO[71]. However, it is not always clear how to link one feature to a specific BP control mechanism. Kurtosis (the feature with the highest importance for RF and 2nd highest for LASSO) for example, represents changes in the overall shape of the PPG. Additionally, there were large variations in the feature importances observed between folds (see figure 5). This may be one contributing factor to the large IQR for RMSE and MAE found for all models using the PPG, ECG and PPG+ECG feature sets. This is especially pertinent given the fact that our study consisted of only a single perturbation of BP induced by an infusion of vasoactive medication.

A secondary explanation for the improved performance of the RF model relative to the LASSO+OLS model, may be that the features presented in this study (or at least the high-performing features presented in figure 5 (a)) have a non-linear relationship to ΔBP. This is further corroborated by the relationship between the top 9 ranking (non-demographic) features and ΔSBP shown in figure 6. Only σ_{g1} shows a discernible global relationship to ΔSBP. Individual, paticipant-specific clusters are apparent, highlighting the low intra-paticipant variability but high inter-paticipant variability. We additionally note that even within the clusters, non-linear relationships are often observed. This significantly impacts the ability to develop *individual calibration* models using these features. For PAT, typically only 2 model parameters (a slope and an intercept) are required for accurate *individual calibration*. Therefore, theoretically, a dataset containing only two measurements of PAT and BP is required for accurate calibration to an individual (although in practice this number is much higher for accurate estimation of the model parameters, see[1]). Whereas, for non-linear modelling of the PPG or ECG features, many more parameters are required to be estimated thus forcing a much larger dataset requirement for accurate *individual calibration*.

It should be noted that many of the features used in this study were derived from fiducial points such as the dicrotic notch. Relying on fiducial point detection has a number of limitations for BP estimation. The detection algorithms often set arbitrary decisions or thresholds for fiducial point locations. As discussed previously, the typical definition for the systolic peak was not appropriate in this study due to the increasing influence of the tidal wave. Additionally, fiducial point detection algorithms will be valid up to a precision; small changes in BP observed in, for example, an ambulatory setting may result in very small perturbations in feature values that are indistinguishable from errors in fiducial point detection[20]. Finally, the fiducial points are not always detectable. The dicrotic notch has been reported to diminish in elderly individuals due to atherosclerosis (hardening of vessel walls and recruitment of collagen fibres to support walls)[43,73]. We found that the majority of the features of high importance (for example: kurtosis, skewness and all Gaussian decomposition features) did not require fiducial point detection. For the reasons stated above, these may offer more desirable representation of changes of the PPG.

Limitations

There are several limitations to the work presented in this study. Firstly, the clinical study evaluated in this work involved perturbing BP via infusion of phenylephrine, an α_1-adrenergic receptor agonist that causes arterial and venous vasoconstriction. In daily life, BP changes are induced by a diverse set of physiological mechanisms governed by the autonomic nervous system. The conclusions presented here must be placed within the context of the mechanisms impacted by the infusion of phenylephrine, namely an increase in TPR. Additionally, our study presents a single perturbation of BP for each participant with BP increased from baseline and then returned to its original value. As highlighted by Mukkamala *et al.*[7], single perturbation studies may lead to biasedly favourable results. It is important that future works evaluate changes of BP across a range of different BP perturbations in individuals.

Secondly, our results were reported across a small number of, relatively homogeneous, healthy participants (N=26). We employed a *hybrid calibration* strategy to estimate changes in BP and utilised information from all available participants via a LOSOCV framework. However, this data-driven strategy requires more individuals to improve model accuracy. This was particularly highlighted in the calibration issue for the four individuals with the largest errors (see Supplementary Information figure SI 3).

Thirdly, despite participants being administered a significant dose of phenylephrine (2mcg/kg/min), the changes we observed in the PPG (see figure 1) were very subtle. We were able to detect the fiducial points of the PPG accurately, as motion artefacts were reduced and the contact pressure of the pulse oximeter was maintained constant. However, in a real-world setting where such large variations in BP are uncommon and motion artefacts are a significant source of noise for PPG, this may be a significant limitation to BP estimation using PPG.

Finally, measurements of BP using a sphygmomanometer cuff are susceptible to various forms of noise that can distort the readings. The oscillometric device used as a BP reference in this study was compliant with the IEC 60601-2-30/EN60601-2-30 and with the American National Standard for Electronic or Automated Sphygmomanometers (ANSI/AAMI SP 10/92)[74] with a maximum mean error of ±5 mmHg (±0.7kPa) and a maximum standard deviation of 8mmHg (1.1kPa). The accuracy of the blood pressure cuff is a significant limitation to using single-point or *hybrid calibration* for BP estimation. Slight errors in a single cuff reading, caused by instrumentation error as well as user error (movement, wrong cuff size, etc) may translate into a consistent offset in BP estimation. Consider, for example, 006 in Supplementary Information figure SI 3, the initial calibration during the rest period sets, with both ΔSBP cuff and ΔSBP est at 0 mmHg. In the following 5 cuff inflations, the ΔSBP cuff readings decreased to just under -5mmHg, within the resolution of the ANSI/AAMI protocol. It is unclear whether this change in SBP is a real change (potentially caused by the participant relaxing after the start of the study) or if it was a result of instrumental errors in the blood pressure cuff. Either way, a consistent DC offset of 5-10mmHg was observed for the remaining BP estimates

in this individual.

5 Conclusion

Under an infusion of phenyleprhine, changes in the PPG (to a greater extent) and the ECG (to a lesser extent) reflect changes in BP that can be tracked using certain morphological features. For monitoring of BP by a single device, we recommend focusing on the PPG as this appears to be far superior to BP monitoring than using the ECG. In this study, we observed clear changes in the PPG in response to the dose increase of phenylephrine and characterised these by smooth muscle activation and a clear increase in the amplitude of the reflected tidal wave. These changes were mirrored in certain features and it appears that their relationship to ΔBP may be non-linear. BP estimation using the PPG may offer similar performance to PAT which has significant limitations as it requires two synchronous devices (ECG and PPG) for accurate measurements. In general, the calibration protocol for accurate BP estimation requires more attention, especially if the relationship is non-linear. *Hybrid calibration* strategies may not adequately reflect the unique and individualised relationship between changes in BP and changes in the PPG. Therefore, they should be used with caution and only as a potential indicator of relative changes as opposed to a clinical assessment of BP.

Data availability

The datasets generated or analysed during the current study are not publicly available due to the sensitive and identifiable nature of our data, patient consent and restrictions of the ethics protocol to protect the privacy of patients involved in the study. Contact eoin.finnegan@eng.ox.ac.uk for any queries.

References

1. Finnegan, E. *et al.* Pulse arrival time as a surrogate of blood pressure. *Sci. reports* **11**, 1–21 (2021).

2. Payne, R. A., Symeonides, C. N., Webb, D. J. & Maxwell, S. R. Pulse transit time measured from the ECG: An unreliable marker of beat-to-beat blood pressure. *J. Appl. Physiol.* **100**, 136–141, DOI: 10.1152/japplphysiol.00657.2005 (2006).

3. Mukkamala, R. *et al.* Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice. *IEEE Transactions on Biomed. Eng.* **62**, 1879–1901, DOI: 10.1109/TBME.2015.2441951 (2015).

4. Sharma, M. *et al.* Cuff-Less and Continuous Blood Pressure Monitoring: A Methodological Review. *Technologies* **5**, 21, DOI: 10.3390/technologies5020021 (2017).

5. Peter, L., Noury, N. & Cerny, M. A review of methods for non-invasive and continuous blood pressure monitoring: Pulse transit time method is promising? *Irbm* **35**, 271–282, DOI: 10.1016/j.irbm.2014.07.002 (2014).

6. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **28**, DOI: 10.1088/0967-3334/28/3/R01 (2007).

7. Mukkamala, R., Hahn, J.-O. & Chandrasekhar, A. Photoplethysmography in noninvasive blood pressure monitoring. In *Photoplethysmography*, chap. 11, 359–400 (Elsevier, 2022).

8. Baruch, M. C. *et al.* Pulse decomposition analysis of the digital arterial pulse during hemorrhage simulation. *Nonlinear biomedical physics* **5**, 1–15 (2011).

9. Vybornova, A. *et al.* 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 Press. Monit.* **30**, 0 (2021).

10. Radha, M. *et al.* Estimating blood pressure trends and the nocturnal dip from photoplethysmography. *Physiol. Meas.* **40**, 025006, DOI: 10.1088/1361-6579/ab030e (2019). 1805.09121.

11. Hosanee, M. *et al.* Cuffless Single-Site Photoplethysmography for Blood Pressure Monitoring. *J. Clin. Medicine* **9**, 723, DOI: 10.3390/jcm9030723 (2020).

12. Elgendi, M. *et al.* The use of photoplethysmography for assessing hypertension. *npj Digit. Medicine* **2**, 60, DOI: 10.1038/s41746-019-0136-7 (2019).

13. Millasseau, S. C., Kelly, R. P., Ritter, J. M. & Chowienczyk, P. J. Determination of age-related increases in large artery stiffness by digital pulse contour analysis. *Clin. Sci.* **103**, 371–377, DOI: 10.1042/cs1030371 (2002).

14. Dehghanojamahalleh, S. & Kaya, M. Sex-Related Differences in Photoplethysmography Signals Measured From Finger and Toe. *IEEE J. Transl. Eng. Heal. Medicine* **7**, 1–7 (2019).

15. Boonya-Ananta, T. *et al.* Synthetic photoplethysmography (PPG) of the radial artery through parallelized Monte Carlo and its correlation to body mass index (BMI). *Sci. reports* **11**, 1–11 (2021).

16. Sun, S., Bezemer, R., Long, X., Muehlsteff, J. & Aarts, R. M. Systolic blood pressure estimation using PPG and ECG during physical exercise. *Physiol. Meas.* **37**, 2154–2169, DOI: 10.1088/0967-3334/37/12/2154 (2016).

17. Miao, F. *et al.* A novel continuous blood pressure estimation approach based on data mining techniques. *IEEE J. Biomed. Heal. Informatics* **21**, 1730–1740, DOI: 10.1109/JBHI.2017.2691715 (2017).

18. Wong, M. Y. M., Poon, C. C. Y. & Zhang, Y. T. An evaluation of the cuffless blood pressure estimation based on pulse transit time technique: A half year study on normotensive subjects. *Cardiovasc. Eng.* **9**, 32–38, DOI: 10.1007/s10558-009-9070-7 (2009).

19. Mukkamala, R. & Hahn, J. O. Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Predictions on Maximum Calibration Period and Acceptable Error Limits. *IEEE Transactions on Biomed. Eng.* **65**, 1410–1420, DOI: 10.1109/TBME.2017.2756018 (2018).

20. Hasanzadeh, N., Ahmadi, M. M. & Mohammadzade, H. Blood Pressure Estimation Using Photoplethysmogram Signal and Its Morphological Features. *IEEE Sensors J.* **20**, 4300–4310 (2019).

21. Slapničar, G., Mlakar, N. & Luštrek, M. Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors* **19**, 3420 (2019).

22. Schlesinger, O., Vigderhouse, N., Moshe, Y. & Eytan, D. Estimation and Tracking of Blood Pressure Using Routinely Acquired Photoplethysmographic Signals and Deep Neural Networks. *Critical Care Explor.* **2**, e0095, DOI: 10.1097/cce.0000000000000095 (2020).

23. Reisner, A. T., Clifford, G. D. & Mark, R. G. The physiological basis of the electrocardiogram. *Adv. methods tools for ECG data analysis* **1**, 25 (2006).

24. Timmermann, V. *et al.* An integrative appraisal of mechano-electric feedback mechanisms in the heart. *Prog. biophysics molecular biology* **130**, 404–417 (2017).

25. Reed, A., Kohl, P. & Peyronnet, R. Molecular candidates for cardiac stretch-activated ion channels. *Glob. Cardiol. Sci. Pract.* **2014**, 19 (2014).

26. Simjanoska, M., Gjoreski, M., Gams, M. & Bogdanova, A. M. Non-invasive blood pressure estimation from ECG using machine learning techniques. *Sensors (Switzerland)* **18**, 1160, DOI: 10.3390/s18041160 (2018).

27. Harford, M. *et al.* Study protocol for an exploratory interventional study investigating the feasibility of video-based non-contact physiological monitoring in healthy volunteers by Mapping Of Lower Limb skIn pErfusion (MOLLIE). *BMJ Open* **10**, e036235 (2020).

28. Richards, E. & V. Maani, C. Phenylephrine. In *StatPearls [Internet]* (StatPearls Publishing, 2019).

29. NICE. Information for the public: High blood pressure (2011).

30. Villarroel, M. *et al.* Non-contact vital-sign monitoring of patients undergoing haemodialysis treatment. *Sci. reports* **10**, 1–21 (2020).

31. Charlton, P. H. *et al.* Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review. *IEEE reviews biomedical engineering* **11**, 2–20 (2017).

32. Chandrasekhar, A., Yavarimanesh, M., Natarajan, K., Hahn, J.-O. & Mukkamala, R. PPG sensor contact pressure should be taken into account for cuff-less blood pressure measurement. *IEEE Transactions on Biomed. Eng.* **67**, 3134–3140 (2020).

33. Orfanidis, S. J. *Introduction to signal processing* (Prentice-Hall, Inc., 1995).

34. Elgendi, M., Liang, Y. & Ward, R. Toward Generating More Diagnostic Features from Photoplethysmogram Waveforms. *Diseases* **6**, 20, DOI: 10.3390/diseases6010020 (2018).

35. Charlton, P. H., Celka, P., Farukh, B., Chowienczyk, P. & Alastruey, J. Assessing mental stress from the photoplethysmogram: A numerical study. *Physiol. Meas.* **39**, DOI: 10.1088/1361-6579/aabe6a (2018).

36. Padilla, J. M. *et al.* Assessment of relationships between blood pressure, pulse wave velocity and digital volume pulse. In *Computers in Cardiology*, vol. 33, 893–896 (IEEE, 2006).

37. Elgendi, M. On the Analysis of Fingertip Photoplethysmogram Signals. *Curr. Cardiol. Rev.* **8**, 14–25, DOI: 10.2174/157340312801215782 (2012).

38. Mok Ahn, J. New aging index using signal features of both photoplethysmograms and acceleration plethysmograms. *Healthc. Informatics Res.* **23**, 53–59, DOI: 10.4258/hir.2017.23.1.53 (2017).

39. Addison, P. S. Slope transit time (STT): A pulse transit time proxy requiring only a single signal fiducial point. *IEEE Transactions on Biomed. Eng.* **63**, 2441–2444 (2016).

40. Lyu, Y. *et al.* Measuring photoplethysmogram-based stress-induced vascular response index to assess cognitive load and stress. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 857–866 (2015).

41. Wang, L., Pickwell-MacPherson, E., Liang, Y. P. & Zhang, Y. T. Noninvasive cardiac output estimation using a novel photoplethysmogram index. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, 1746–1749, DOI: 10.1109/IEMBS.2009.5333091 (IEEE, 2009).

42. Shin, H. & Min, S. D. Feasibility study for the non-invasive blood pressure estimation based on ppg morphology: Normotensive subject study. *BioMedical Eng. Online* **16**, DOI: 10.1186/s12938-016-0302-y (2017).

43. Takazawa, K. *et al.* Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform. *Hypertension* **32**, 365–370, DOI: 10.1161/01.HYP.32.2.365 (1998).

44. Pilt, K., Meigas, K., Ferenets, R., Temitski, K. & Viigimaa, M. Photoplethysmographic signal waveform index for detection of increased arterial stiffness. *Physiol. Meas.* **35**, 2027–2036, DOI: 10.1088/0967-3334/35/10/2027 (2014).

45. Rubins, U., Grabovskis, A., Grube, J. & Kukulis, I. Photoplethysmography analysis of artery properties in patients with cardiovascular diseases. In *IFMBE Proceedings*, vol. 20 IFMBE, 319–322, DOI: 10.1007/978-3-540-69367-3-85 (Springer, 2008).

46. Couceiro, R. *et al.* Multi-Gaussian fitting for the assessment of left ventricular ejection time from the Photoplethysmogram. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 3951–3954, DOI: 10.1109/EMBC.2012.6346831 (IEEE, 2012).

47. Xing, X. *et al.* Robust blood pressure estimation from finger photoplethysmography using age-dependent linear models. *Physiol. measurement* **41**, 25007 (2020).

48. Solà, J. & Delgado-Gonzalo, R. *The Handbook of Cuffless Blood Pressure Monitoring: A Practical Guide for Clinicians, Researchers, and Engineers* (Springer Nature, 2019).

49. Wang, L., Xu, L., Feng, S., Meng, M. Q. & Wang, K. Multi-Gaussian fitting for pulse waveform using Weighted Least Squares and multi-criteria decision making method. *Comput. Biol. Medicine* **43**, 1661–1672, DOI: 10.1016/j.compbiomed.2013.08.004 (2013).

50. Abdi, H. & Williams, L. J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2**, 433–459, DOI: 10.1002/wics.101 (2010).

51. Pan, J. & Tompkins, W. J. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomed. Eng.* **BME-32**, 230–236, DOI: 10.1109/TBME.1985.325532 (1985).

52. Li, Q., Mark, R. G. & Clifford, G. D. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiol. Meas.* **29**, 15–32, DOI: 10.1088/0967-3334/29/1/002 (2008).

53. Yang, S. *et al.* Blood pressure estimation with complexity features from electrocardiogram and photoplethysmogram signals. *Opt. Quantum Electron.* **52**, 135, DOI: 10.1007/s11082-020-2260-7 (2020).

54. Leonarduzzi, R. F., Schlotthauer, G. & Torres, M. E. Wavelet leader based multifractal analysis of heart rate variability during myocardial ischaemia. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 110–113 (IEEE, 2010).

55. Li, H. *et al.* A new ECG signal classification based on WPD and ApEn feature extraction. *Circuits, Syst. Signal Process.* **35**, 339–352 (2016).

56. Higuchi, T. Approach to an irregular time series on the basis of the fractal theory. *Phys. D: Nonlinear Phenom.* **31**, 277–283 (1988).

57. Ogedegbe, G., Agyemang, C. & Ravenell, J. E. Masked hypertension: Evidence of the need to treat. *Curr. Hypertens. Reports* **12**, 349–355, DOI: 10.1007/s11906-010-0140-4 (2010).

58. Lakhal, K., Ehrmann, S. & Boulain, T. Noninvasive BP Monitoring in the Critically Ill: Time to Abandon the Arterial Catheter? *Chest* **153**, 1023–1039, DOI: 10.1016/j.chest.2017.10.030 (2018).

59. Pollock, D. S. G. Smoothing with cubic splines (1993).

60. Mukkamala, R., Stergiou, G. S. & Avolio, A. P. Cuffless blood pressure measurement. *Annu. Rev. Biomed. Eng.* **24** (2022).

61. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).

62. Natarajan, K. *et al.* Photoplethysmography Fast Upstroke Time Intervals Can Be Useful Features for Cuff-Less Measurement of Blood Pressure Changes in Humans. *IEEE Transactions on Biomed. Eng.* (2021).

63. Dormann, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).

64. Kim, J. H. Multicollinearity and misleading statistical results. *Korean journal anesthesiology* **72**, 558 (2019).

65. Genuer, R., Poggi, J.-M. & Tuleau, C. Random Forests: some methodological insights. *arXiv preprint arXiv:0811.3619* (2008).

66. Taji, B., Shirmohammadi, S., Groza, V. & Batkin, I. Impact of skin–electrode interface on electrocardiogram measurements using conductive textile electrodes. *IEEE Transactions on Instrumentation Meas.* **63**, 1412–1422 (2013).

67. Von Wowern, E., Östling, G., Nilsson, P. M. & Olofsson, P. Digital photoplethysmography for assessment of arterial stiffness: Repeatability and comparison with applanation tonometry. *PLoS ONE* **10**, e0135659, DOI: 10.1371/journal.pone.0135659 (2015).

68. Laurent, S. *et al.* Expert consensus document on arterial stiffness: methodological issues and clinical applications. *Eur. heart journal* **27**, 2588–2605 (2006).

69. Vincent, J. L. Understanding cardiac output. *Critical Care* **12**, DOI: 10.1186/cc6975 (2008).

70. Bennis, F. C. *et al.* Artifacts in pulse transit time measurements using standard patient monitoring equipment. *PLoS ONE* **14**, e0218784, DOI: 10.1371/journal.pone.0218784 (2019).

71. Lin, W.-H., Li, X., Li, Y., Li, G. & Chen, F. Investigating the physiological mechanisms of the photoplethysmogram features for blood pressure estimation. *Physiol. measurement* **41**, 44003 (2020).

72. Cannesson, M., Jian, Z., Chen, G., Vu, T. Q. & Hatib, F. Effects of phenylephrine on cardiac output and venous return depend on the position of the heart on the Frank-Starling relationship. *J. applied physiology* **113**, 281–289 (2012).

73. Millasseau, S. C., Ritter, J. M., Takazawa, K. & Chowienczyk, P. J. Contour analysis of the photoplethysmographic pulse measured at the finger. *J. hypertension* **24**, 1449–1456 (2006).

74. Stergiou, G. S. *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* **71**, 368–374, DOI: 10.1161/HYPERTENSIONAHA.117.10237 (2018).

Acknowledgements

Author contributions

Study design and conceptualisation was performed by MH, PW, LT and MV. Data collection was performed by MH, EF, SD, and MV. EF, MV, SD, and MH developed methodology and software. MV, LT, and PW provided supervision. EF prepared the first draft of this manuscript. All authors critiqued and edited the manuscript for intellectual content.

Competing interest

LT and PW report significant grants from the National Institute of Health Research (NIHR), UK and the NIHR Biomedical Research Centre, Oxford, during the conduct of the study; modest grants and personal fees from Sensyne Health, outside the submitted work. EF, SD, MH, and MV declare no competing interests.

Additional information

Correspondence

All requests for material should be addressed to EF.