Hypothesis Scoring for Confidence-Aware Blood Pressure Estimation with Particle Filters

Jonathan Martinez, *Student Member, IEEE*, Bryant Passage, *Student Member, IEEE,* Bobak J. Mortazavi, Senior *Member, IEEE,* and Roozbeh Jafari, Senior *Member, IEEE*

***Abstract*— We propose our Confidence-Aware Particle Filter (CAPF) framework that analyzes a series of estimated changes in blood pressure (BP) to provide several true state hypotheses for a given instance. Particularly, our novel confidence-awareness mechanism assigns likelihood scores to each hypothesis in an effort to discard potentially erroneous measurements – based on the agreement amongst a series of estimated changes and the physiological plausibility when considering DBP/SBP pairs. The particle filter formulation (or sequential Monte Carlo method) can jointly consider the hypotheses and their probabilities over time to provide a stable trend of estimated BP measurements. In this study, we evaluate BP trend estimation from an emerging bio-impedance (Bio-Z) prototype wearable modality although it is applicable to all types of physiological modalities. Each subject in the evaluation cohort underwent a hand-gripper exercise, a cold pressor test, and a recovery state to increase the variation to the captured BP ranges. Experiments show that CAPF yields superior continuous pulse pressure (PP), diastolic blood pressure (DBP), and systolic blood pressure (SBP) estimation performance compared to ten baseline approaches. Furthermore, CAPF performs on track to comply with AAMI and BHS standards for achieving a performance classification of Grade A, with mean error accuracies of -0.16 ± 3.75 mmHg for PP (r=0.81), 0.42 ± 4.39 mmHg for DBP (r=0.92), and -0.09 ± 6.51 mmHg for SBP (r=0.92) from more than test 3500 data points.**

***Index Terms*—Blood pressure, blood pressure variability, particle filtering, confidence-aware, ensemble modeling**

I. Introduction

Cardiovascular diseases (CVDs) remain amongst the leading causes for death worldwide [1]. Of the risk factors associated with CVDs, high blood pressure (BP) has been identified as the most prominent [2]. While a majority of the population measure BP levels only during annual medical examinations, ambulatory BP monitoring (ABPM) provide continuous measurements throughout patients' daily activities [3], generating comprehensive insight for healthcare providers and thus enabling the early detection of life-threatening illnesses. The current gold standard ABPM devices depend on repeated cuff inflations that cause discomfort over time. However, advances in non-invasive wearable devices – leveraging modalities such as bio-impedance (Bio-Z) [4], photoplethysmogram (PPG) [5], or electrocardiogram (ECG) [6] – enable a more convenient alternative while providing beat-to-beat BP monitoring. Yet, existing theoretical modeling and machine learning-based solutions focus on time-invariant absolute estimation that only considers physiological waveform morphology information from the current cardiac cycle or time window for which a measurement will be assigned. Furthermore, they do not consider any of their own previous or future decision making, thus do not incorporate any notion of self-correction when estimates deviate from the expected trend. Therefore, as these conventional approaches are susceptible to occasionally generating inaccurate BP estimates due to data- and model-dependent error [7]–[9], a single or series of erroneous measurements could potentially lead to misdiagnosis by healthcare providers. In this study, we propose a novel alternative for achieving beat-to-beat BP estimation that derives the absolute value trend from a series of estimated changes. By tracking our own estimated instances over time, we generate multiple BP hypotheses for the current cardiac cycle and probabilistically identify the true state through our novel confidence-awareness mechanism.

Traditional non-invasive BP estimation with physiological waveforms obtained from wearable modalities such as Bio-Z, PPG, and ECG involves the theoretical modeling of blood flow. Popular blood flow measurements include pulse transit time (PTT), pulse arrival time (PAT), and pulse wave velocity (PWV), which measure the time and force by which blood travels from the heart through the arteries [10]. The complete theoretical model which enables derivation of absolute BP requires consideration of additional arterial characteristics [11] – such as elasticity, diameter, and length, which may only be precisely obtained invasively or with additional sensors. There are also patient- and population-specific parameters which must be manually calibrated for by domain experts.

Data-driven solutions are ideal alternatives to theoretical modeling that removes the burden associated with any manual

J. Martinez is with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX, 77840 USA (e-mail: jmartinez0304@tamu.edu)

B. Passage is with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX, 77840 USA (e-mail: bryant.passage@tamu.edu)

B. Mortazavi is with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX, 77840 USA (e-mail: bobakm@tamu.edu)

R. Jafari is with the Department of Electrical and Computer Engineering, School of Engineering Medicine, and Computer Science and Engineering, Texas A&M University, College Station, TX, 77840 USA (e-mail: rjafari@tamu.edu)

calibration or obtaining arterial characteristics. Such approaches leverage artificial intelligence and machine learning – such as with decision trees, support vector machines, or neural network models – to automatedly identify the relationship between captured waveforms' discriminative features and their corresponding target label through an iterative analysis of sample data [12]. However, their precision is dependent on obtaining abundant training instances that adequately represent all possible variations of the input-to-output parings that may exist. Otherwise, challenges associated with imbalanced representation of target labels in training distributions [8], [9], the potential presence of noisy waveform instances input to the regression models (in both training and testing) [13], and data shift due to contextual variation over time [14] could lead to estimation error. Moreover, existing beat-to-beat BP estimation approaches are time-invariant by only considering waveform morphology for the current cardiac cycle (or time-window) and do not leverage any prior information from previous decision-making instances. Therefore, erroneous measurements for a single or for a series of cardiac cycles without any ability to self-correct will lead to inaccurate trends.

The straightforward solution to model self-correction is frequent retraining (or recalibration) with new instances of the physiological waveform that have evolved either over time due to fluctuating health states or upon new contexts (i.e., various activities of daily living) [15]. Yet, aggregating additional training data manually by domain experts is burdensome and will occur infrequently such that a misdiagnosis may have already occurred with the incorrectly reported BP measurements. Other approaches that attempt to ensure stable BP estimation incorporate Kalman filters [16] and particle filters [17], which can achieve unsupervised state-space modeling. Typically, such approaches track blood flow characteristics (e.g., PAT, morphological characteristics, etc.) over time and attempt to leverage signal quality indices to determine the amount of trend correction that should be applied. However, non-linear modeling of stochastic processes with Kalman filters depend on local linear projections approximated by Taylor series expansion, therefore previous work has shown more precise state-space modeling by particle filters which leverage sequential Monte Carlo sampling [18]. Furthermore, particle filters are better equipped to handle non-Gaussian noise and may more seamlessly handle dynamic sets of hypotheses.

We propose our Confidence-Aware Particle Filter (CAPF) as a novel alternative BP estimation approach that incorporates comparisons between collected physiological waveform morphologies associated with a current cardiac cycle to several of those preceding it in an effort to estimate changes in BP – providing several BP hypotheses. Furthermore, we introduce a novel particle weight update mechanism that establishes confidence to each candidate absolute BP value derived from the estimated changes to mitigate potentially erroneous estimates – namely Agreement Scoring and Multi-Tasked Scoring. Through this, we can extract the most likely true state hypotheses to yield high-quality absolute pulse pressure (PP), diastolic blood pressure (DBP), and systolic blood pressure (SBP) measurements – where PP is defined as the difference between SBP and DBP. This trend information constructed from the series of changes can be fused with the aforementioned time-invariant estimates to produce a more robust ultimate BP measurement. We demonstrate CAPF improved continuous BP estimation performance for an emerging Bio-Z wearable modality, although the proposed framework may be applied to all types of physiological modalities. The contributions of this work are as follows:

- We propose a novel approach for constructing beat-to-beat BP trends through frequently estimated BP changes when comparing the waveform morphological features between two independent cardiac cycles
- We propose our Confidence-Aware Particle Filter framework that probabilistically manages multiple BP hypotheses through the novel Agreement Scoring and Multi-Tasked Scoring particle weight update mechanisms
- We demonstrate improved continuous BP estimation by CAPF when compared to ten baseline approaches, while analyzing an emerging wearable modality such as Bio-Z

II. Related Work

In this research, we pursue beat-to-beat analysis where conventional data-driven approaches typically rely on first extracting waveform morphology features that would reflect physiological responses. Such features include PAT, PTT, PWV, and a number of time-, amplitude-, and area-based features [10]. They are then input to a variety of machine learning models to directly estimate absolute BP values for the cardiac cycle waveform (or time window of waveforms) being analyzed. The most popular shallow models for cuffless BP estimation – leveraging modalities such as ECG, PPG, and Bio-Z – include Lasso regression [19], support vector machines (SVM) [20], k-nearest neighbors (KNN) [21], and XGBoost [22]. Even amidst limited training datasets, they provide interpretable decision making based on feature similarity. Yet, they tend to struggle in the presence of data shift. Alternatively, deep learning models – such as multilayer perceptrons (MLP) [23], convolutional neural networks (CNN) [24], and recurrent neural networks (RNN) [25] – provide more complex, non-linear modeling. However, their performance is dependent on abundant training data with sufficient instance variation.

State-space modeling of a series of observations to ensure stable trend generation have proven successful with methods such as Kalman and particle filtering [26]. Kalman filters attempt to model the underlying processes of state transitions and therefore have been proposed in previous work to track both waveform features, signal quality indices (SQIs), and target BP behavior over time [27]. However, Kalman filters assume the state-transitioning to be linear and that noise will be Gaussian. Alternatively, particle filtering is a probabilistic, Monte Carlo algorithm that is capable of simultaneously managing multiple partial observations to model a time-series [28]. Appealing characteristics of this solution include the tracking of a posterior distribution, the ability to model short-term sequences, and the robustness to instances of noisy observations. Particle filters have been proposed for physiological sensing tasks such as waveform denoising [29] or physiological parameter estimation [17], [30]. However, the particle filter formulations for previous

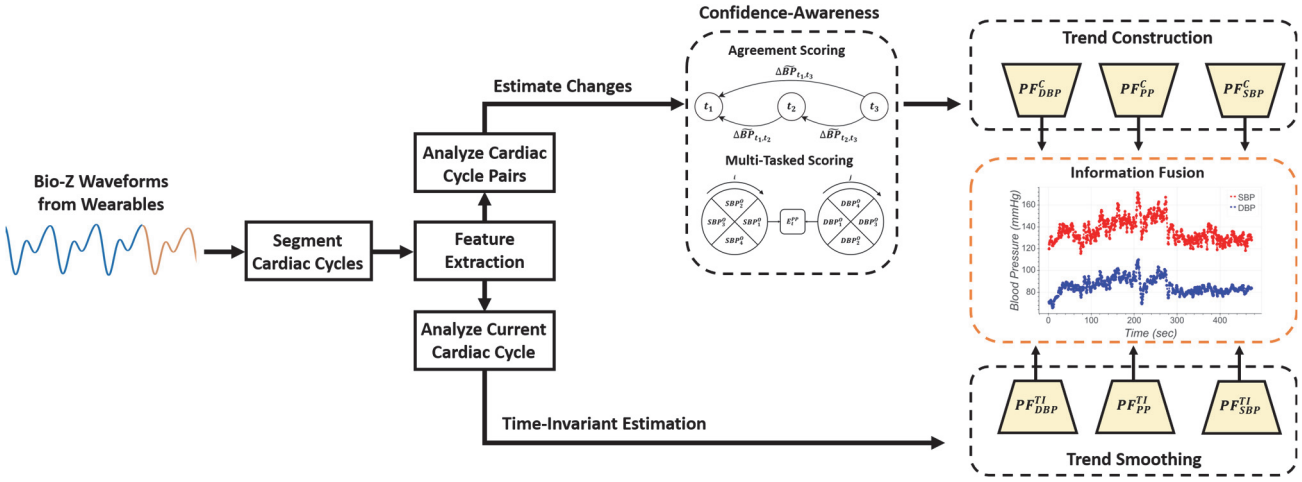

Fig. 1. Overview of the proposed Confidence-Aware Particle Filter (CAPF) framework that achieves beat-to-beat BP estimation by leveraging both information from trends constructed by a series of estimated BP changes and time-invariant information.

work generally rely on an implicit linear relationship between observations' characteristics and the task which is being pursued. Unfortunately, BP possesses a complex non-linear relationship with existing cuffless wearable sensor modalities.

III. Methodology

The proposed CAPF framework – shown in Fig. 1 – achieves robust, real-time BP estimation. Provided only the initial ground truth PP, DBP, and SBP absolute values obtained in the calibration phase before any testing, we construct BP trends with a following series of estimated beat-to-beat changes. Therefore, for the desired BP estimate associated with any given cardiac cycle, we are granted several hypotheses for our final measurement provided by each of the comparisons to all of the preceding cardiac cycles, in addition to the time-invariant estimates. To extract a notion of high-quality consensus from them, CAPF mitigates erroneous measurements by assigning a confidence score to each hypothesis driven by our novel Agreement Scoring and Multi-Tasked Scoring particle weight update mechanisms. In this study, we focus our analysis on the emerging Bio-Z wearable modality and evaluate BP estimation performance of CAPF when compared to ten baseline approaches – although our framework may be applied to all types of physiological modalities. In the following subsections, we first discuss estimating changes in BP, then we present the formulation for particle filtering, and finally we discuss how final BP estimates are produced while quantified confidence scores help to mitigate erroneous measurements.

A. Estimating Changes in BP with Bio-Z

Bio-Z has demonstrated potential in previous work as a desirable alternative wearable modality to existing cuff-based ABPM devices as it may be captured at the wrist with a more comfortable-to-wear configuration and extracts higher resolution beat-to-beat information [4]. This non-invasive electrical signal is obtained by injecting a current into the wrist and capturing the voltage difference between two electrodes placed along the radial artery. The 2-channel waveform morphology corresponds to the changes in blood volume during the systolic and diastolic phases (along with the reflection pressure), which are represented by the characteristic peak, notch, and max slope characteristic fiducial points as shown in Fig. 2. Particularly, the diastolic peaks (DIA1, DIA2) indicate the start/end of a cardiac cycle and the minimum pressure when the heart is expanding to be re-filled with blood. The systolic notch (SYS) represents the maximum pressure when the heart contracts to pump blood throughout the body, where the inflection point (IP) is the reflection pressure. Thus, the max slope (MS) point grants insight into the rate of change between the two prominent cardiac phases. These fiducial points may be identified with the physiological waveforms' first and second derivative morphological characteristics such as peak, valley, and zero-crossing points. From the identified fiducial points for each cardiac cycle, we can derive PTT, inter-beat interval (IBI), time-, amplitude-, and area-based features that reflect the vascular properties of the radial artery and may be mapped to each BP type – PP, DBP, and SBP. PTT is the time distance between the MS point locations from each Bio-Z channel. The IBI feature is extracted for each channel as the time distance between the MS points for a current and the immediately following cardiac cycle. The time-, amplitude-, and area-based features are extracted by comparing the DIA1 point to the MS, SYS, and IP point locations for each channel. This yields 23 features extracted for a given cardiac cycle.

When applying existing time-invariant estimation solutions

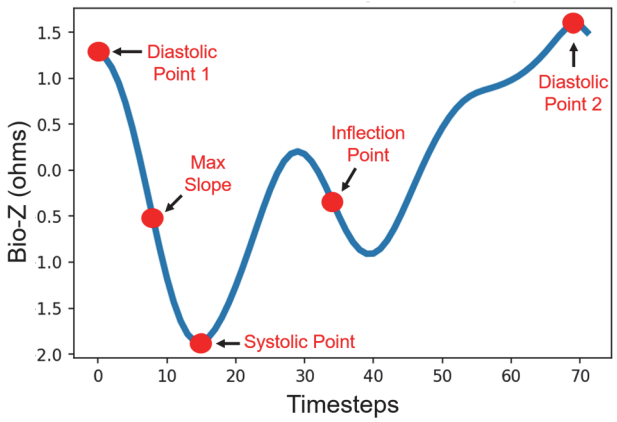

Fig. 2. Sample Bio-Z waveform morphology for a single cardiac cycle with labeled fiducial points. Here, timesteps represents each sample of the waveform kept arbitrary in the interest of visualization.

to beat-to-beat BP estimation, the derived physiological waveform features for a single cardiac cycle are directly input to the estimation model. Alternatively, with our proposed approach, when estimating changes in BP between two independent cardiac cycles we will concatenate the extracted features for each into one set, and jointly input them into the model to estimate the change between the absolute BP values that would be associated with each. Particularly, in testing, the estimation operation may be formulated as

$$\Delta\widetilde{BP}_{t-d,t} = f(X_{t-d}, X_t) \tag{1}$$

where X_t is the feature vector associated with the current cardiac cycle, while X_{t-d} is the feature vector associated with an independent cardiac cycle that existed d seconds in the past, and $\Delta\widetilde{BP}_{t-d,t}$ is the estimated change in BP between the two cardiac cycles. Furthermore, for any given $\Delta\widetilde{BP}_{t-d,t}$, a change is always defined as the absolute BP value corresponding to the cardiac cycle at time t minus that associated with the cardiac cycle from time $t-d$, where the estimated change could be either an increase or decrease. In this way, the corresponding absolute BP value measurement may be derived from the estimated change as

$$\widetilde{BP}^{o}_{t-d,t} = \widetilde{BP}_{t-d} + \Delta\widetilde{BP}_{t-d,t} \tag{2}$$

where we add $\Delta\widetilde{BP}_{t-d,t}$ to the previously estimated absolute BP value (generated by the end-to-end CAPF framework) associated with a cardiac cycle from the past, $\widetilde{BP}_{t-d}$, to obtain a hypothesis BP measurement for the current cardiac cycle (treated as observations in particle filter formulation), $\widetilde{BP}^{o}_{t-d,t}$ – where the subscript pairing of "$t-d,t$" indicates an approximated absolute BP value derived from an estimated change. This proposed approach allows direct comparison between each of the physiological waveform morphologies. In this study for our CAPF framework, we use XGBoost [22] regression models for both time-invariant and change estimation as its efficacy has been demonstrated in previous work that have pursued the standard absolute BP estimation problem setup – although, our proposed framework is formulated as model-agnostic. Furthermore, separate XGBoost models will be trained to estimate changes for each BP type (i.e., PP, DBP, and SBP independently).

When training the change in BP estimation model, we create sample instances by concatenating all possible cardiac cycle feature pairs for the available training data. However, the primary challenge is managing physiological waveform morphological variance with respect to context [14], [31]. That is, since patient health state and activity is dynamic over time, the waveform morphology's relationship to BP may also evolve. Therefore, when constructing training sample pairs, we constrain the maximum time distances between two independent cardiac cycles that are being compared to be 60 seconds in an effort to increase the likelihood that each pair exists in the same context. However, since intuitively we understand that there will always exist an exponentially higher number of training instance pairs with shorter time distances, in testing we reduce the constraint to 30 seconds. This is since a majority of training sample pairs will have time distances under 30 seconds, yet, our testing estimation performance may still benefit from the exposure in training to potentially larger changes in BP – which are typically associated with sample pairs with time distances greater than 30 seconds but less than 60 seconds. In Section IV Experiments and Results we will present an ablation study that demonstrates the impact of increasing this maximum time distance threshold.

B. Particle Filter Formulation

As we obtain a series of estimated BP changes, while jointly considering them with the time-invariant information, CAPF will probabilistically translate them into continuous, beat-to-beat absolute BP value measurements. Given only the initial ground truth PP, DBP, and SBP measurements acquired through the calibration phase, we incrementally construct the following trend as future cardiac cycles occur. Furthermore, for any given estimated instance, a variable number of BP hypotheses will be considered from the preceding cardiac cycles (within the 30 seconds). Straightforward approaches would typically apply a variation of averaging for finding consensus amongst multiple hypotheses [32]. However, when a given (or multiple) erroneous hypothesis is present, these straightforward approaches will yield inaccurate measurements as they do not incorporate any notion of uncertainty to mitigate such error. On the contrary, particle filtering treats these hypotheses as observations and incorporates a scoring mechanism that could be used to describe the reliability of each candidate absolute BP value estimate [17]. This is achieved by tracking a set of sampled particles (or set of candidate BP values). Initially, particles are randomly sampled to represent the whole range of plausible BP values, and, as new information is obtained from each cardiac cycle, particle likelihoods are updated to represent the posterior distribution that models the state-space of BP trends. Furthermore, by resampling from the whole posterior distribution after every update, the particle filter is able to retain prior information obtained from the preceding BP trend and leverage it to provide a more robust BP estimate. In this subsection, we will first introduce our general formulation for how particle filtering manages the observations and their corresponding confidence scores as shown in Fig. 3 (consistent with previous work also pursuing physiological parameter estimation [33]); then, in the following subsection (Section III-C) we will further discuss how the confidence scores are generated.

First, before any estimation, the particle filter is initialized with a prior distribution representing the range of plausible BP values that may be observed in the near future,

$$BP_t \sim \pi_X = U(BP_{min}, BP_{max}) \tag{3}$$

where π_X represents a uniform prior distribution such that all values between the physiologically plausible BP_{min} and BP_{max} range have an equal likelihood of being the next true state – before any observations are provided. From this, a pre-defined number of particles, N_p, are sampled from π_X to represent the dynamic posterior distribution that will be updated over time with each new set of observations. When particles are initially sampled, each particle value is assigned a uniform weight that represent the likelihood of being the subsequent true BP value

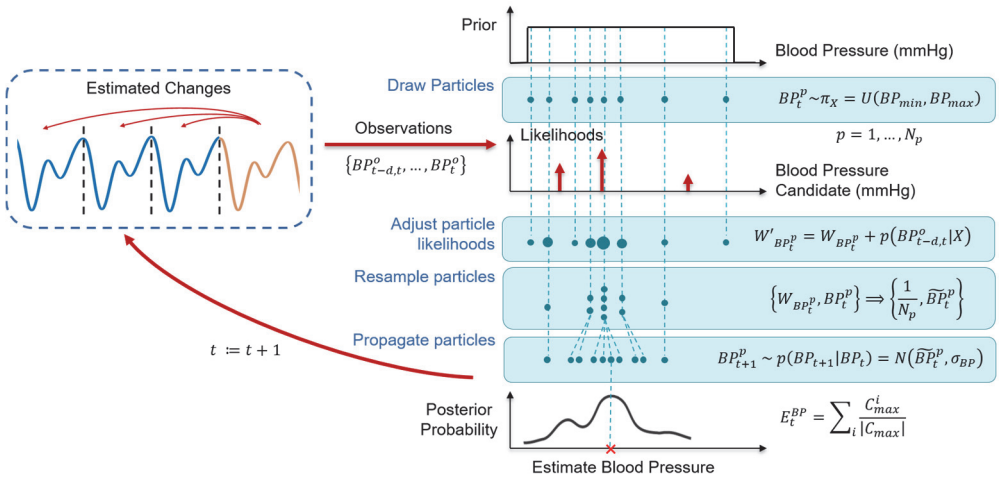

Fig. 3. Particles are sampled from a prior distribution to represent the dynamic posterior distribution of a state-space by tracking the likelihoods of candidate true state values – which are updated according to observations and our confidence-awareness mechanism. Then, particles are resampled and propagated, where the final estimate is identified as the candidate BP value with the largest number of surrounding particles.

$$W_{BP_t^p} = \frac{1}{N_p} \qquad (4)$$
$$\forall p \in (1, N_p)$$

where BP_t^p is the p^{th} particle value at any given time t, and $W_{BP_t^P}$ is the associated likelihood for the p^{th} particle at time t.

As observations are provided, particle weights will be augmented to reflect the appropriate likelihood of each candidate absolute BP value being the true state according to the new evidence. For now, we will denote this probability for an observation as $p(BP^o_{t-d,t} \mid X)$ in an effort to more clearly move forward with presenting the particle filter formulation – where X is a vector of the input waveform morphological features. To be explained in the following subsection, these likelihoods are effectively defined as the confidence scores. Nevertheless, these observation probabilities will augment the likelihood of each particle whose BP value corresponds to it

$$W'_{BP_t^p} = W_{BP_t^p} + p(BP^o_{t-d,t} \mid X) \qquad (5)$$
$$\forall p \in (1, N_p)$$

where $W'_{BP_t^p}$ is now the updated particle weight. Thus, executing an update to the posterior distribution from which we can now resample particles from it according to the new likelihoods. However, when hypotheses are provided with high confidence scores, it is likely that particle weights may become severely skewed. Therefore, we employ the commonly used sequential importance resampling (SIR) to alleviate any particle degeneracy for those representing BP hypotheses from previous estimated instances [34]

$$M_t^p = \sum_{r=1}^{p} \widehat{W}_{BP_t^r}, \forall p \in (1, N_p) \qquad (6)$$

$$u = \underset{a}{argmin} \Big| R_u \sim U(0,1) \le M_t^a \qquad (7)$$

$$BP'^p_t = BP^u_t, \forall p \in (1, N_p) \qquad (8)$$

where M_t^p is an element of the cumulative sum vector calculated over the particle weights, R_u is a randomly sampled number between 0 and 1, and BP'^p_t is a resampled particle state.

To yield the final estimate, we first cluster particles centered on all candidate BP values within a range, CS, of 3 mmHg,

$$C_N \triangleq Set\ of\ all\ BP'^p_t \Big| \left|BP'^m_t - BP'^n_t\right| < CS$$
$$\forall m \in (1, N_p), \forall n \in (1, N_p)$$
$$E_t^{BP} = \sum_i \frac{C^i_{max}}{|C_{max}|} \qquad (9)$$

where C_N is a cluster of particles centered on a candidate BP value, C^i_{max} is an element of the largest cluster, and E_t^{BP} is the absolute BP value by which the largest cluster was centered – regarded as the final estimation for this instance.

Last, in preparation for subsequent estimates, each particle value is randomly shifted to reflect plausible changes in BP which are expected to be relatively small for consecutive beats and possibly larger for beats that are farther apart in time. This shift function is defined as

$$BP^p_{t+1} \sim f(BP^p_t) \sim N(BP'^p_i, \sigma_{BP}) = BP'^p_i + (\sigma_{BP} \times R_N \sim N(0,1))$$
$$\forall p \in (1, N_p) \qquad (10)$$

where R_N is a randomly sampled number from a normal distribution to be applied as a shift factor, and σ_{BP} is the expected change in BP between estimated instances.

Then, we will move on to the next cardiac cycle and repeat this process from equation (5) until the end of the captured physiological waveform stream. Through this series of updates to the posterior distribution over time, the particle filter will inherently track the temporal characteristics of the constructed BP trends.

The above formulation of the particle filter is designed for estimation of one type of BP trend at a time – in other words, the PP, DBP, and SBP trends constructed from the series of

changes will be independently tracked by separate particle filters. In the following subsection, we will explain how we obtain $p\left(BP^{o}_{t-d,t} \mid X\right)$ for each estimated change.

C. Confidence-Aware Particle Filter for BP Estimation

Determining the appropriate confidence level that should be assigned to each observation is the most critical mechanism of the particle filter formulation and is the defining characteristic of our proposed CAPF framework, as this operation is responsible for ultimately extracting consensus among the multiple hypotheses and ensuring the final BP estimate is high-quality. We robustly assign probabilities to each observation through two scoring types: 1) Agreement Scoring and 2) Multi-Tasked Scoring. Ultimately, a likelihood score will be generated for each and combined to provide $p\left(BP^{o}_{t-d,t} \mid X\right)$ – moreover, an effective particle weight augmentation.

Agreement Scoring mechanism checks for agreement amongst a series of estimated changes in BP. Using the example in Fig. 4, for any given set of three consecutive cardiac cycles (e.g., existing in time at t_1, t_2, and t_3) there are 3 changes in BP that take place: $\Delta\widetilde{BP}_{t_1,t_2}$, $\Delta\widetilde{BP}_{t_2,t_3}$, and $\Delta\widetilde{BP}_{t_1,t_3}$. These estimates can be further categorized as those between cardiac cycles that are consecutive ($\Delta\widetilde{BP}_{t_1,t_2}$, $\Delta\widetilde{BP}_{t_2,t_3}$) and non-consecutive ($\Delta\widetilde{BP}_{t_1,t_3}$). In the ideal case where we assume each of these are accurate, we expect a non-consecutive estimate to be the sum of all the consecutive estimates that exist between its independent cardiac cycles that are being compared

$$\Delta\widetilde{BP}_{t_1,t_3} = \Delta\widetilde{BP}_{t_1,t_2} + \Delta\widetilde{BP}_{t_2,t_3} \tag{11}$$

On the contrary, if any of the three estimated changes is inaccurate, the above property will not hold. Therefore, we can leverage this phenomenon to achieve Agreement Scoring by measuring the absolute difference between the non-consecutive estimate and the sum of the consecutive estimates

$$A_{t_1,t_3} = \left|\Delta\widetilde{BP}_{t_1,t_3} - (\Delta\widetilde{BP}_{t_1,t_2} + \Delta\widetilde{BP}_{t_2,t_3})\right| \tag{12}$$

To relate this concept back to probability assignment for particle filters, while referring back to Fig. 4 we can see that the comparisons executed for a given estimated instance may all be categorized as non-consecutive except when estimating change in BP from the immediately preceding cardiac cycle. Therefore, we may compute an $A_{t-d,t}$ agreement score/difference for each of them and SoftMax normalize them into $W^{A}_{t-d,t}$ scores such that the observation that yields the smallest difference will receive a maximum score of $\frac{1}{N_p}$ (which is the unit weight for a sampled particle) and that the largest difference will receive the minimum score of $\frac{0.05}{N_p}$ (5% of the unit weight), while all others will be scaled to this range. Last, the single consecutive change in BP estimate will receive a score of $\frac{0.5}{N_p}$ to reflect random chance since it could not be evaluated with this property.

On the other hand, Multi-Tasked Scoring ensures physiological plausibility of reported PP, DBP, and SBP values. As aforementioned, PP is measured as SBP minus DBP where it is conceptually defined as the force by which blood is flowing

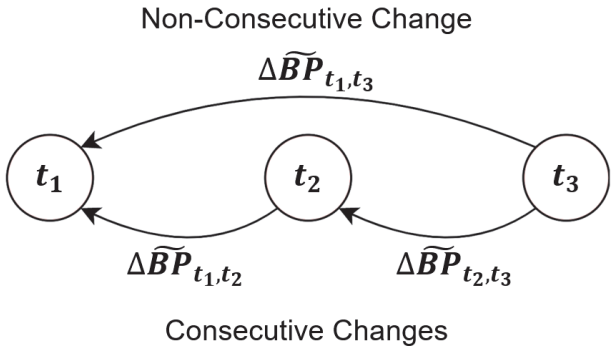

Fig. 4. Example series of cardiac cycles to visualize non-consecutive and consecutive changes used in Agreement Scoring.

due to pressure exerted by the heart's contractions and inherent arterial compliance [35]. Recall that estimated PP, DBP, and SBP trends constructed by the series of estimated changes are tracked independently by distinct particle filters (i.e., PF^{C}_{PP}, PF^{C}_{DBP}, PF^{C}_{SBP}). Therefore, we can use the estimated PP value produced by PF^{C}_{PP} to check for plausibility between SBP and DBP observations to further enhance the confidence assignment mechanism. This may be formulated as

$$\omega^{MT}_{i} \mathrel{+}= \begin{cases} 1, & SBP^{o}_{i} \approx DBP^{o}_{j} + \overline{PP}_{t} \mid \forall DBP^{o}_{j} \in (1,J) \\ 0, & otherwise \end{cases} \tag{13}$$

$$\omega^{MT}_{j} \mathrel{+}= \begin{cases} 1, & DBP^{o}_{j} \approx SBP^{o}_{i} - \overline{PP}_{t} \mid \forall SBP^{o}_{i} \in (1,I) \\ 0, & otherwise \end{cases} \tag{14}$$

Where SBP^{o}_{i} represents an SBP observation while there are I of them in total, and DBP^{o}_{j} represents a DBP observation while there are J of them in total. Therefore, ω^{MT}_{i} is a counter for each SBP^{o}_{i} observation where we iterate through the whole set of DBP^{o}_{j} observations to identify the number of plausible counterparts according to the final E^{PP}_{t} provided by PF^{C}_{PP} (within an empirically determined tolerance of 0.5 mmHg to account for inherent noise) – as visualized in Fig. 5. We can repeat this same process in vice versa for each DBP^{o}_{j} observation and its ω^{MT}_{j} counter.

Last, to update the particle weight augmentation step from Section III-B, $p\left(BP^{o}_{t-d,t} \mid X\right)$ is redefined as the multiplication between the likelihoods obtained with Agreement Scoring and Multi-Tasked Scoring

$$W'_{BP^{p}_{t}} = W_{BP^{p}_{t}} + p\left(BP^{o}_{t-d,t} \mid X\right) \tag{15}$$

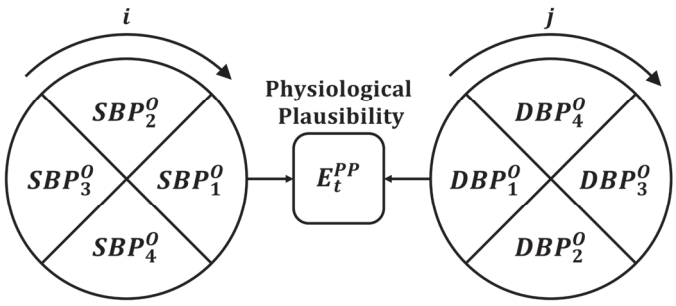

Fig. 5. Conceptual model for the Multi-Tasked Scoring mechanism where each set of SBP and DBP observation pairings will be evaluated for physiological plausibility based on the estimated PP value.

$$p\left(BP^{o}_{t-d,t}\middle|X\right) = W^{A}_{t-d,t} * \omega^{MT}_{t-d,t} \quad (16)$$

where $W^{A}_{t-d,t}$ is the likelihood obtained by Agreement Scoring and $\omega^{MT}_{t-d,t}$ is a likelihood obtained by Multi-Tasked Scoring for an observation (whether SBP or DBP). Through this formulation, $\omega^{MT}_{t-d,t}$ may remove any contribution for an observation that has no physiologically plausible counterparts, thus discarding the risk for erroneous measurements.

Now that we have constructed the estimated BP trend using the series of changes, we will re-incorporate the time-invariant estimation. Using only the waveform morphology features for the current cardiac cycle, X_t, the time-invariant model will directly estimate an absolute BP value. However, as mentioned earlier in this paper, with the constrained amount of information available during this approach, it is more likely that we may encounter erroneous measurements. To alleviate this, we leverage our presented particle filter formulation to inherently model the temporal characteristics of each BP types' trend estimated by the time-invariant model – where each particle filter can be referred to as PF^{TI}_{PP}, PF^{TI}_{DBP}, and PF^{TI}_{SBP}. Although, since for this setting there will only be one absolute BP hypothesis associated with each cardiac cycle, particle weight updates will be incremented with a unit value ($\frac{1}{N_p}$). This will provide us with a smoothed BP trend constructed with the time-invariant information. Finally, we produce the ultimate absolute BP value for each type as the average between the E^{BP}_t values provided by PF^{C}_{BP} and PF^{TI}_{BP}. Fusing the information from each enables self-correction for each particle filter. The time-invariant information will stabilize the BP trend constructed with the series of changes from error propagation, and similarly the several hypotheses provided by comparisons to preceding cardiac cycles will highlight the true beat-to-beat variation.

IV. Experiments and Results

A. Bio-Z for Physiological Sensing and Data Collection

The Bio-Z dataset analyzed for this study was collected under the IRB approval by IRB2020-0090F at Texas A&M University where informed consent was obtained for each subject. Throughout each data collection trial, Bio-Z was captured by a prototype wrist-worn device introduced in previous work [4] and simultaneously reference BP waveforms were continuously measured from the Finapres NOVA System as each subject underwent a hand-gripper exercise, a cold pressor test, and a recovery state to simulate the contextual variability that typical wearable devices may encounter with activities of daily living. The Finapres NOVA System (https://www.finapres.com/) leverages a finger cuff to generate arterial waveforms that are calibrated with a single brachial cuff-based measurement. Each 8-minute trial was captured and stored into independent 4-minute data streams, and repeated four times in the same day to encompass a collection session (32 minutes in total). Bio-Z wrist-worn prototype electrode sensors are aligned with the radial arteries to monitor blood flow. Therefore, in addition to human physiological variation, waveform morphology is also impacted by electrode sensor alignment quality with the arteries, contact with the skin, and motion. To ensure a fair evaluation, we discarded any cardiac cycle waveforms corrupted by noise – identified with a heuristic-based denoising filter that discarded instances that do not follow the expected pulse wave morphologies. In addition, we applied a 10-second rolling window to smooth both waveform features and reference BP values in an effort to alleviate the inherent abovementioned noise (which is consistent with the previous work that originally proposed the utility of Bio-Z [4]). We also ensured that each subject included for analysis obtained collection sessions whose trials were composed of at least 50% clean waveforms – thus providing eight subjects for analysis. The subjects in this cohort were between the ages of 18 to 40, consisting of four males and four females. Of this cohort, four subjects contained BMI values within the healthy range (21.6 to 23.9), one subject is within the underweight range (14.8), one subject is within the overweight range (26.8), and two subjects are within the obese range (31.5 and 35.7). With respect to the BP values captured in this dataset, the distribution of DBP have a range from 50.3 to 116.7 mmHg with a mean value of 81.6 mmHg and a standard deviation of 12.0 mmHg, while the SBP distribution has a range from 87.7 to 175.3 mmHg with a mean value of 130.3 mmHg and a standard deviation of 15.5 mmHg.

B. Personalized BP Estimation Performance

We evaluate CAPF's performance for personalized PP, DBP, and SBP estimation compared to ten baseline approaches. The first set of baseline methods (five models) achieve conventional time-invariant beat-to-beat BP estimation, where only the extracted features from the waveform associated with the cardiac cycle are input to each regression model for direct estimation of the corresponding absolute BP value. We included the existing state-of-the-art regression models for this scenario: XGBoost [22], support vector machine (SVM) with radial basis function kernel [20], k-nearest neighbors (KNN) [21], Lasso [19], and multi-layer perceptron (MLP) [23] consisting of three fully connected layers with rectified linear unit (ReLU) activation with 100, 50, and 1 hidden units respectively. The second baseline method type (one model) extracts a series features from cardiac cycles within a 5-second time window to be analyzed by a recurrent neural network (Window-RNN) [25] consisting of a bi-directional gated recurrent units (GRU) layer with 128 hidden units, followed by a fully connected layer of 100 hidden units, and concluded with a final output fully connected layer. This approach takes a step towards expanding input information through the notion of sequential modeling, yet still conducts time-invariant decision-making. The third set of baseline methods (two models) uses XGBoost to estimated changes in BP over time to construct the absolute value trend – again taking another step towards leveraging prior information but this time also considering previous estimates to introduce the notion of generating multiple hypotheses for the true state of the current instance. The first model consists of averaging all hypotheses to identify the final estimate, solely relying on consensus. The second model uses a vanilla particle filter without our proposed confidence-awareness mechanism to introduce the impact of state-space tracking in addition to finding consensus. Finally, the last set of baseline methods (two models) serve as an ablation study for the scoring the approaches presented in Section III-C. First, we evaluate CAPF if we were to only leverage Agreement Scoring (CAPF-AS); then, we evaluate

TABLE I
PERSONALIZED BP ESTIMATION PERFORMANCE

Model	Pulse Pressure			Diastolic Blood Pressure			Systolic Blood Pressure		
	$RMSE$ $(mmHg)$	$ME \pm SD$ $(mmHg)$	r	$RMSE$ $(mmHg)$	$ME \pm SD$ $(mmHg)$	r	$RMSE$ $(mmHg)$	$ME \pm SD$ $(mmHg)$	r
XGBoost	4.25	-0.68 ± 4.19	0.760	6.91	-2.04 ± 6.60	0.824	9.52	-2.80 ± 9.10	0.834
SVM	4.13	0.82 ± 4.04	0.778	7.11	-2.15 ± 6.79	0.810	8.91	-1.22 ± 8.22	0.847
KNN	4.08	0.39 ± 4.06	0.789	7.55	-2.03 ± 7.27	0.777	9.73	-1.46 ± 9.62	0.810
Lasso	5.27	1.74 ± 4.98	0.656	7.02	-2.34 ± 6.62	0.821	9.39	-1.13 ± 9.32	0.826
MLP	6.43	-0.33 ± 6.42	0.607	10.91	1.16 ± 10.8	0.603	12.24	0.27 ± 12.2	0.732
Window-RNN	4.31	0.45 ± 4.28	0.770	8.39	-2.53 ± 8.00	0.731	10.08	-1.95 ± 9.88	0.799
Ensemble Average	5.21	0.16 ± 5.20	0.713	7.85	2.68 ± 7.37	0.848	8.69	2.16 ± 8.41	0.877
Vanilla PF	4.89	0.34 ± 4.58	0.735	7.58	2.51 ± 7.45	0.860	8.33	1.78 ± 8.42	0.880
CAPF-AS	5.66	0.49 ± 5.64	0.653	6.91	-1.24 ± 6.80	0.820	11.03	-0.70 ± 11.0	0.768
CAPF-MT	5.63	0.56 ± 5.61	0.657	7.16	-1.11 ± 7.07	0.800	10.52	-0.14 ± 10.5	0.796
CAPF (Our Method)	**3.75**	**-0.16 ± 3.75**	**0.813**	**4.41**	**0.42 ± 4.39**	**0.925**	**6.51**	**-0.09 ± 6.51**	**0.918**

TABLE II
COMPARISON OF METHODS TO BHS STANDARDS

Model	Diastolic Blood Pressure				Systolic Blood Pressure			
	% ≤ **5 $mmHg$**	% ≤ **10 $mmHg$**	% ≤ **15 $mmHg$**	$Grade$	% ≤ **5 $mmHg$**	% ≤ **10 $mmHg$**	% ≤ **15 $mmHg$**	$Grade$
XGBoost	59	86	95	B	43	73	87	C
SVM	58	84	95	B	49	76	90	C
KNN	57	83	93	B	46	74	87	C
Lasso	55	86	95	B	46	74	90	C
MLP	45	71	85	C	40	65	80	C
Window-RNN	55	80	90	B	48	73	85	C
Ensemble Average	50	82	96	B	48	78	93	C
Vanilla PF	51	81	96	B	50	78	93	B
CAPF-AS	55	84	98	B	44	68	80	C
CAPF-MT	56	84	96	B	42	68	83	C
CAPF (Our Method)	**78**	**97**	**99**	**A**	**61**	**89**	**97**	**A**

CAPF if we were to only leverage Multi-Task Scoring (CAPF-MT). This study will analyze the impact of each mechanism.

For personalized BP estimation, each model was evaluated with a given subjects' data for training and testing. To ensure a fair evaluation, for each subject, the trial (composed of both independent 4-minute streams) with the largest number of cardiac cycles available was designated as the test set to be completely excluded from any training. Therefore, we executed a k-fold cross validation process with grid search only using the remaining trials designated for training to identify all model hyperparameters. XGBoost hyperparameters set for CAPF were optimized for the max tree depth and number of estimators – where the range of values were [2, 3, 4, 5, 6, 7, 8, 9, 10] and [25, 50, 75, 100] respectively. The hyperparameter set to be retrained with all training data only was determined as the one which achieved estimation performance closest to the mean across all folds – with respect to the root mean square error (RMSE) and correlation. Across all subjects, the most reported hyperparameter set was a max depth of 3 with 100 estimators. SVM was optimized for the degree of its polynomial function and its regularization parameter – where the range values were [2,3,4,5] and [0.1,1,10] respectively. KNN was optimized for the number of nearest neighbors parameter – where the range values were from 1 to 30. Lasso was optimized for its shrinkage factor – where the range values were [0.1,1,10]. Both MLP and Window-RNN were optimized with a learning rate of 0.001, batch size of 32, and were trained for a maximum of 1000 epochs where 10% of the training set was tracked as validation to determine early stopping. Last, for fairness, all parameters for the vanilla particle filter and CAPF were consistently fixed across all subjects (as formulated in Section III) with the number of particles sampled (N_p) set to 1000.

In Table I, we compare the performance for personalized absolute PP, DBP, and SBP estimation with each model using RMSE, Mean Error (ME) plus and minus the standard deviation of errors (±SD), and the Pearson correlation coefficient between reference and estimated BP trends (r), calculated with the test data for all subjects. Our proposed CAPF framework achieves superior performance compared to each of the baselines for all three BP types with respect to all metrics, while MLP was consistently the poorest performer (due to its sensitivity to imbalanced training datasets). For SBP, in general, constructing absolute value trends from estimated changes showed improvements to all of the time-invariant methods as they leverage the least amount of information for analysis. For PP and DBP, we observed the reverse outcome. Moreover, for each BP type, we observe a marginal improvement when applying Vanilla PF to multiple hypotheses compared to the straightforward averaging approach – reflecting the need for probabilistic state-space modeling. On the other hand, the ablation study analyzing the impact of each scoring mechanism independently (CAPF-AS and CAPF-MT) reflects their codependence. Particularly, both variants' estimation performance is amongst the poorest performers for PP and SBP while they are among the average performers for DBP.

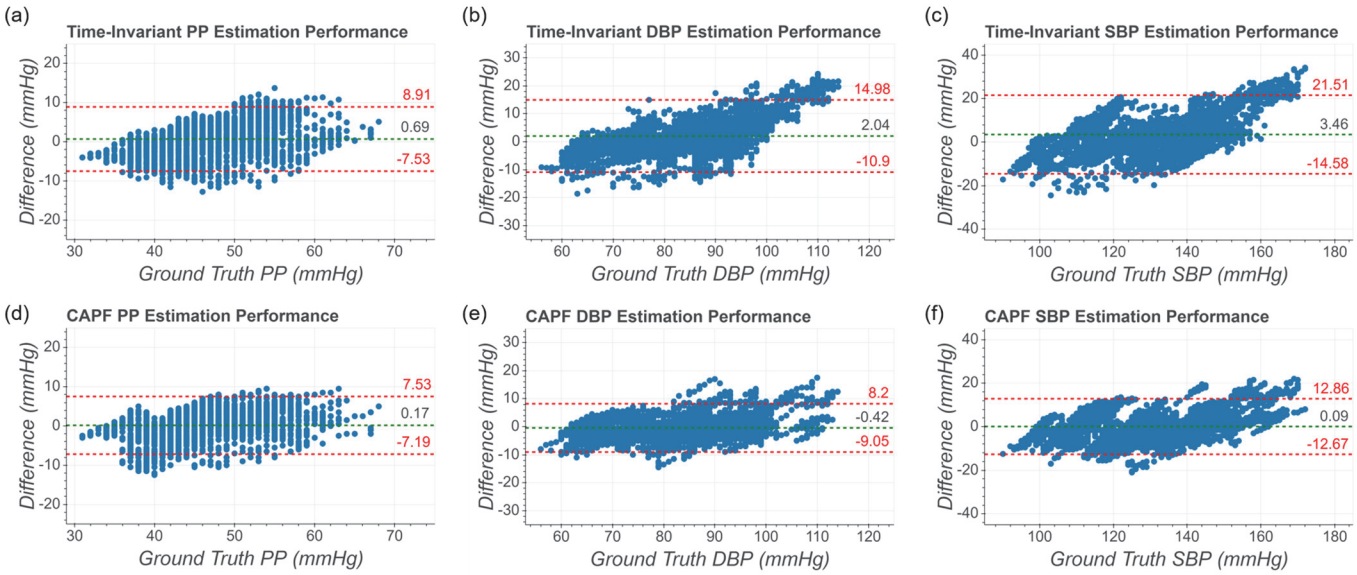

Fig. 6. Bland-Altman difference plots used to compare the time-invariant model (a,b,c) to CAPF (d,e,f) for PP, DBP, and SBP value estimation, where difference is defined as the ground truth minus the estimate. It should be noted that all estimated BP values were reported as integers.

Therefore, without the integration of the two scoring mechanisms to compensate for each other, each variant will bias its decision making to introspection and physiological plausibility respectively. Last, the largest performance improvement achieved by our CAPF framework for each BP type demonstrates the robustness provided when introducing the multiple hypotheses (from both change and time-invariant estimation) and our novel confidence-awareness mechanism for the formulated particle filter.

We also compare BP performance for all methods to the Association for the Advancement of Medical Instrumentation (AAMI) and the British Hypertension Society (BHS) standards [36]. The AAMI standards require a ME ≤ 5 mmHg and a SD ≤ 8 mmHg, where shown in Table I all evaluated methods achieve this for DBP except for MLP, yet only our proposed CAPF framework achieves this for SBP. The BHS standards require that of the reported absolute differences, at least 60% should be ≤ 5 mmHg, 85% should be ≤ 10 mmHg, and 95% should be ≤ 15 mmHg to achieve a Grade A classification. However, it should be noted that these standards typically require a minimum of 85 subjects for evaluation with at most 3 individual BP measurements for each, which yields 255 test data points in total. Yet, since our analysis targeted continuous BP monitoring, we captured several more cardiac cycles for each subject to yield over 3500 test data points for our evaluation. Therefore, we choose to compare model performance to these standards to remain consistent with previous work and since the sample size of our evaluation is significantly larger than their minimum requirement. As indicated in Table II, our proposed CAPF framework is the only evaluated method which is on track to achieve a Grade A classification for more than 3500 test data points – for both DBP and SBP. For DBP, all baseline methods are on track to achieve a Grade B except for MLP which achieves Grade C. For SBP, all baseline methods are on track to achieve a Grade C except for Vanilla PF which achieves Grade B.

In Fig. 6, we visualize the improved PP, DBP, and SBP estimation performance with our proposed CAPF framework compared to only its time-invariant estimation component (XGBoost without particle filtering) through Bland-Altman Difference plots – it should be noted that all estimated BP values were reported as integers. For DBP and SBP, we observe a reduced margin of error by approximately 31%. For PP, we observe a reduced margin of error by approximately 10%. Particularly, from all plots we identify the main sources of error in the upper and lowermost ranges of BP which are typically underrepresented in training; however, our proposed CAPF framework shows better performance for these regions.

C. Analysis of CAPF Estimation Confidence

An additional benefit of the proposed CAPF framework is our ability to extract a notion of estimation confidence derived from our novel observation weighting mechanism. This characteristic is especially critical when estimating physiological parameters since healthcare providers depend on the reported measurements to diagnose patient health. Otherwise, unnoticed erroneous measurements could prevent the early detection of life-threatening illnesses. Fortunately, our proposed Agreement Scoring and Multi-Tasked Scoring schemes achieve a robust quality check for every hypothesis.

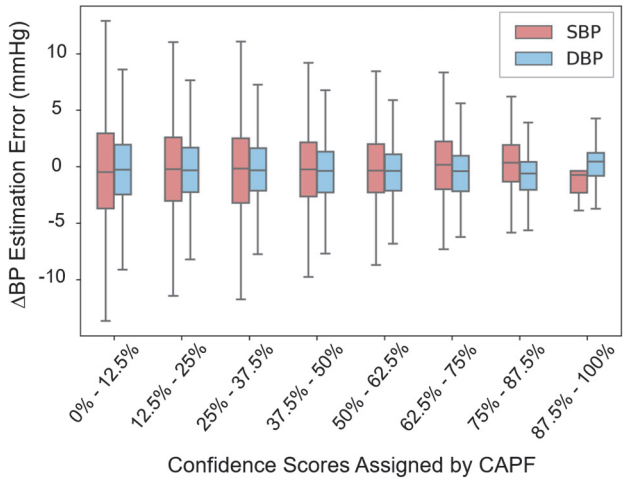

Fig. 7. Analysis on the distribution of estimated BP change errors for DBP and SBP, with respect to the confidence scores assigned by CAPF.

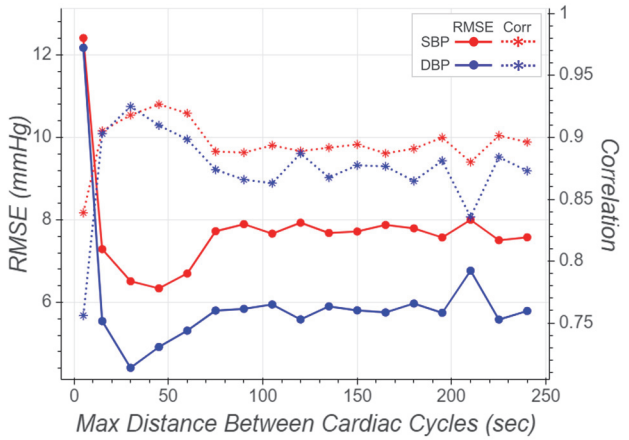

Fig. 8. Change in absolute DBP (blue) and SBP (red) estimation performance with respect to RMSE and Pearson correlation coefficient, as we increase the maximum time distance constraint.

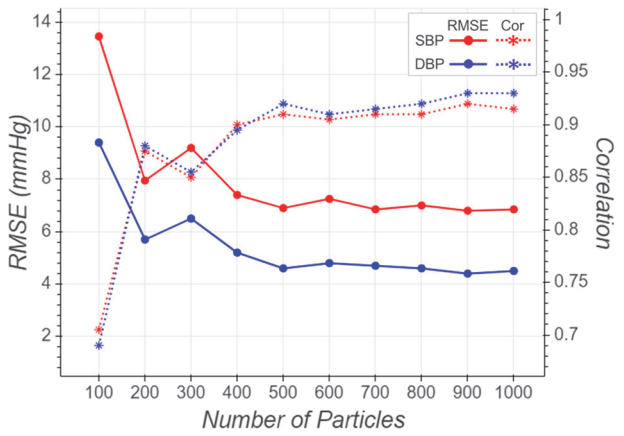

Fig. 9. Change in absolute DBP (blue) and SBP (red) estimation performance with respect to RMSE and Pearson correlation coefficient, as we increase the number of particles used by CAPF.

While we have already observed the improvement to final BP estimation performance upon considering such information in real-time (i.e., Table I), in Fig. 7 we take a more detailed look at how well the assigned observation likelihoods, $p(BP^{o}_{t-d,t}|X)$, predict the quality of estimated changes in SBP and DBP. We plotted the distribution of errors (measured as the reference value minus the estimate) for estimated changes in BP when observations were assigned low confidence scores (where the minimum likelihood is normalized to 0%) through when observations were assigned high confidence scores (where the maximum likelihood is normalized to 100%). From the boxplots for both SBP and DBP, we observe the distribution of errors linearly tighten close to 0 mmHg as the confidence scores approach the maximum score of 100%.

D. Analysis of Maximum Time Distance Between Cardiac Cycles

Although CAPF's particle filtering formulation is agnostic to the type of model used for estimation, a limitation is that we must assume that the estimation models perform reasonably well. One key observation from the framework design phase was that the precision for estimating changes in BP incrementally deteriorates as the time distance between two distinct cardiac cycles becomes farther – as aforementioned, due to contextual variation and that the larger BP changes are more likely to exist for pairs more separated in time. For our core experimental results, we constrained the maximum time distance between cardiac cycles for which we are estimating change to be 30 seconds in testing. However, in Fig. 8, we show the results for an ablation analysis where we incrementally increased the maximum time distance constraint from 15 seconds to 240 seconds (the full length of each independent 4-minute stream). Here, we show that the best performance was achieved when the constraint was set to 30 seconds. As the maximum time distance was increased past this setting, we observe a decline in performance up until 75 seconds where the error stabilizes onward. This reflects how our proposed confidence-awareness framework is capable of identifying the most high-quality BP hypotheses even amongst dramatically increasing the number of hypotheses to be considered by CAPF.

E. Time Complexity Analysis of CAPF

For the vanilla particle filter, time complexity may be expressed as $O(N_p\omega_o)$ where we iterate through each of the sampled particles to execute posterior distribution updates while also searching for the corresponding observation that will be applying the update – such that ω_o is the maximum number of observations that may be available at any given estimated instance. This is since each observation will be increasing particle likelihood with a fixed unit weight ($\frac{1}{N_p}$). For our proposed CAPF framework, recall that we replace this fixed weight with an adaptive likelihood score through Agreement Scoring and Multi-Tasked Scoring. Therefore, the time complexity of only the Agreement Scoring mechanism is $O(\omega_o)$ where we iterate through each of the observations to check for agreement (through a series of mathematical operations in constant time). On the other hand, Multi-Tasked Scoring is expressed as $O(\omega_o^2)$ where we must jointly iterate through the set of observations for both PF_{SBP} and PF_{DBP}. Thus, considering the time complexity of Agreement Scoring and Multi-Tasked Scoring together can be expressed as $O(\omega_o + \omega_o^2)$. Finally, this extends the overall time complexity of CAPF to $O(N_p(\omega_o + \omega_o^2))$. In Fig. 9, we show impact to CAPF performance as we increase N_p from 100 to 1000 by increments of 100. With every incremental increase, the SBP, DBP, and PP performance gradually improves respectively and saturates to our reported RMSE and correlation scores. This accurately depicts that N_p is effectively the resolution by which we can accurately track the posterior distribution.

V. Discussion

Our above experiments demonstrate CAPF improvements to continuous PP, DBP, and SBP estimation compared to ten baseline methods. In Fig. 10 we plot DBP and SBP estimation performance over time for the subject which both the time-invariant model and CAPF performed the best. Our proposed framework reduced RMSE for DBP from 5.70 mmHg to 3.09 mmHg and for SBP from 5.06 mmHg to 3.16 mmHg. An interesting observation for the time-invariant performance is that in most cases if we consider the series of estimates, the short-term fluctuations often resemble that of the ground truth although there are discrepancies with respect to the magnitude of change per cardiac cycle and offsets from the absolute value of the BP type. This is generally a reflection of the sample

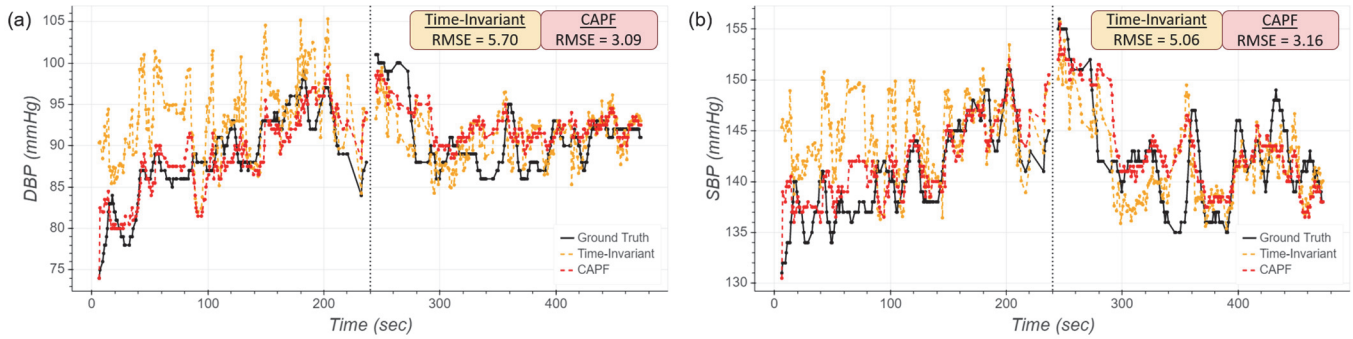

Fig. 10. Estimated beat-to-beat a) DBP and b) SBP values plotted over time, comparing CAPF (red) to the time-invariant model (orange) with respect to ground truth (black) – for both 4-minute data streams tested for one subject (distinguished by the dotted vertical line).

distribution from training, such that values less than 140 mmHg for SBP and less than 85 mmHg for DBP were underrepresented in training. In addition, this could reflect waveform morphology feature shift over time. On the contrary, CAPF does capture these fluctuations more tightly and better follows the true absolute BP values. This indicates that incorporating change in BP estimates enhances generalizability by providing a set of hypotheses that are normalized.

Recall, that CAPF's final BP estimates are extracted from the particle distribution. That is, candidate true state values with the most neighboring particles are interpreted as having the greatest likelihood. Therefore, we demonstrate how the confidence-awareness mechanism scores are translated into particle value distributions. In Fig. 11 we show these distributions for estimated DBP and SBP values for given cardiac cycle instances from the generated trend from Fig. 10. For high-quality estimates when absolute error is 0 mmHg, we observe a unimodal distribution where particles are concentrated around a single peak reflecting strong consensus. Furthermore, the range of candidate BP values spans approximately 20 mmHg. On the contrary, for low-quality estimates, when absolute error is 8 mmHg and 6 mmHg, we observe a multimodal distribution where each peak indicates a candidate BP value with some level of plausibility reflecting poor consensus. Furthermore, the range of candidate BP values increases to approximately 100 mmHg. Despite that CAPF is still able to provide a final estimate for these cases, this reflects a notion of uncertainty.

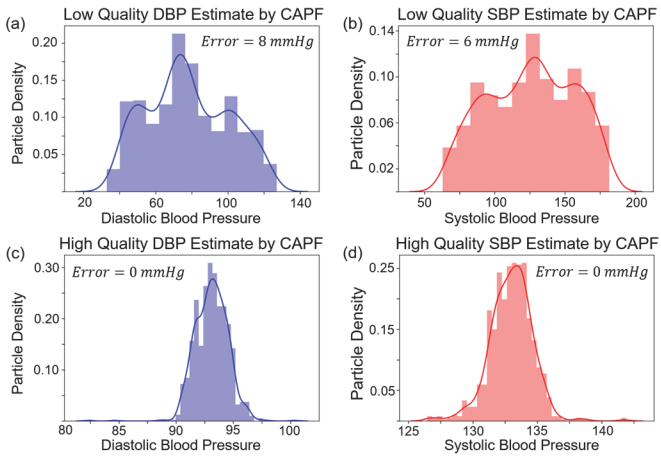

Fig. 11. Particle value distributions for high- and low-quality BP estimates (with respect to absolute error) yielded by CAPF.

A. Limitations

The absolute BP trends constructed from the series of estimated changes require a single initial reference point obtained from calibration (for each independent data stream). In the real-world setting, this reference could be obtained with home cuff-based device or with a robust time-invariant cuffless BP estimator – moreover, healthcare providers may analyze the normalized variability trend [37]. Ideally, the smallest time distance possible between the calibration sample and the first subsequent estimated instance is preferred to increase the likelihood for high precision. Yet, we rely on the particle filter formulation with our incorporated confidence-awareness to prevent the constructed BP trend from wandering due to aggregate error over time. The intention is that this calibration point is effective for the entire length of the collection trial, and this claim was supported by our analysis. In Fig. 10, we also show that CAPF achieves higher-quality beat-to-beat BP estimation and is able to capture all detailed trend fluctuations as BP rises/falls throughout the duration of the tested trial. This indicates that our CAPF framework should not require any additional calibration points for any following estimation. However, a limitation of this study is that the evaluated test trials for each subject is maximum 8-minutes in length. Longer-term round-the-clock collection, such as for the duration of several months, would enable further investigation into the frequency by which recalibration would be required.

VI. Conclusion

In this work, we proposed CAPF for estimating BP trends by leveraging information obtained through frequently estimated beat-to-beat changes and also time-invariant waveform morphology features. CAPF mitigated erroneous measurements with our proposed confidence-awareness mechanism achieved by novel Agreement Scoring and Multi-Tasked Scoring, and showed to be a strong indicator for estimation uncertainty. We evaluated CAPF alongside ten baseline approaches for continuous PP, DBP, and SBP estimation performance when analyzing an emerging wearable modality, Bio-Z. CAPF outperforms all baseline methods with respect to RMSE, ME, and Pearson's correlation coefficient, on track to achieve a Grade A classification performance according to the AAMI/BHS BP estimation standards for more than 3500 test data points. Future opportunities for this work include evaluating over a longer-term round-the-clock collection for durations such as several months. This would enable

investigation into the frequency by which recalibration or model retraining should occur. Moreover, extending the solution to remove the need for obtaining a reference calibration point would strengthen the real-world applicability of the framework. Last, conducting experimentation on framework generalizability would also strengthen its utility.

REFERENCES

[1] A. Timmis *et al.*, "European Society of Cardiology: cardiovascular disease statistics 2021," *Eur. Heart J.*, vol. 43, no. 8, pp. 716–799, 2022.

[2] K. Bogale and A. Aderaw, "Clinical Characteristics and in-Hospital Mortality in Patients with Acute Heart Failure at Dessie Referral Hospital, Northeast Ethiopia," *Abyssinia J. Sci. Technol.*, vol. 6, no. 1, pp. 33–37, 2021.

[3] E. D. Anstey *et al.*, "Diagnosing masked hypertension using ambulatory blood pressure monitoring, home blood pressure monitoring, or both?," *Hypertension*, vol. 72, no. 5, pp. 1200–1207, 2018.

[4] B. Ibrahim and R. Jafari, "Cuffless Blood Pressure Monitoring from an Array of Wrist Bio-impedance Sensors using Subject-Specific Regression Models: Proof of Concept," *IEEE Trans. Biomed. Circuits Syst.*, 2019.

[5] G. Slapni Č Ar, N. Mlakar, and M. Luštrek, "Blood Pressure Estimation from Photoplethysmogram Using a Spectro-Temporal Deep Neural Network," *Sensors (Basel).*, vol. 19, no. 15, Aug. 2019.

[6] S. Yang, Y. Zhang, S. Y. Cho, R. Correia, and S. P. Morgan, "Non-invasive cuff-less blood pressure estimation using a hybrid deep learning model," *Opt. Quantum Electron.*, vol. 53, no. 2, pp. 1–20, Feb. 2021.

[7] Q. Ye, B. Wing-Kuen Ling, N. Xu, Y. Lin, and L. Hu, "Multi-model fusion of classifiers for blood pressure estimation," *IET Syst. Biol.*, vol. 15, no. 6, pp. 184–191, 2021.

[8] J. Martinez, Z. Nowroozilarki, R. Jafari, and B. J. Mortazavi, "Data-Driven Guided Attention for Analysis of Physiological Waveforms With Deep Learning," *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 11, pp. 5482–5493, 2022.

[9] L. N. Guo *et al.*, "Bias In, Bias Out: Underreporting and Underrepresentation of Diverse Skin Types in Machine Learning Research for Skin Cancer Detection--A Scoping Review," *J. Am. Acad. Dermatol.*, 2021.

[10] E. Martinez-R\'\ios, L. Montesinos, M. Alfaro-Ponce, and L. Pecchia, "A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data," *Biomed. Signal Process. Control*, vol. 68, p. 102813, 2021.

[11] X. Ding and Y. T. Zhang, "Pulse transit time technique for cuffless unobtrusive blood pressure measurement: from theory to algorithm," *Biomed. Eng. Lett.*, vol. 9, no. 1, pp. 37–52, Feb. 2019.

[12] S. Maqsood *et al.*, "A survey: From shallow to deep machine learning approaches for blood pressure estimation using biosensors," *Expert Syst. Appl.*, p. 116788, 2022.

[13] Y. Lu, M. Li, B. Wu, Y. Tang, and Z. Wei, "Denoising of pulse wave signal by wavelet packet transform," in *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2021, pp. 232–236.

[14] Q. Yousef, M. B. I. Reaz, and M. A. M. Ali, "The analysis of PPG morphology: investigating the effects of aging on arterial compliance," *Meas. Sci. Rev.*, vol. 12, no. 6, p. 266, 2012.

[15] D. Barvik, M. Cerny, M. Penhaker, and N. Noury, "Noninvasive Continuous Blood Pressure Estimation from Pulse Transit Time: A review of the calibration models," *IEEE Rev. Biomed. Eng.*, 2021.

[16] D. A. Hullender and O. R. Brown, "Simulations of blood pressure and identification of atrial fibrillation and arterial stiffness using an extended Kalman filter with oscillometric pulsation measurements," *Comput. Methods Programs Biomed.*, vol. 198, p. 105768, 2021.

[17] V. Nathan and R. Jafari, "Particle Filtering and Sensor Fusion for Robust Heart Rate Monitoring Using Wearable Sensors," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 6, pp. 1834–1846, 2018.

[18] I. V Stelzer, J. Kager, and C. Herwig, "Comparison of particle filter and extended kalman filter algorithms for monitoring of bioprocesses," in *Computer Aided Chemical Engineering*, vol. 40, Elsevier, 2017, pp. 1483–1488.

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

[20] B. Zhang, H. Ren, G. Huang, Y. Cheng, and C. Hu, "Predicting blood pressure from physiological index data using the SVR algorithm," *BMC Bioinformatics*, vol. 20, pp. 1–15, 2019.

[21] C. Yi, C. Jian, and J. Wenqiang, "Continuous blood pressure measurement based on photoplethysmography," in *2019 14th IEEE International Conference on Electronic Measurement \& Instruments (ICEMI)*, 2019, pp. 1656–1663.

[22] S. Banerjee, B. Kumar, A. P. James, and J. N. Tripathi, "Blood Pressure Estimation from ECG Data Using XGBoost and ANN for Wearable Devices," in *2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2022, pp. 1–4.

[23] M. S. Tanveer and M. K. Hasan, "Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ANN-LSTM network," *Biomed. Signal Process. Control*, vol. 51, pp. 382–392, 2019.

[24] H. Eom *et al.*, "End-To-End Deep Learning Architecture for Continuous Blood Pressure Estimation Using Attention Mechanism," *Sensors (Basel).*, vol. 20, no. 8, Apr. 2020.

[25] C. El Hajj and P. A. Kyriacou, "Cuffless and continuous blood pressure estimation from ppg signals using recurrent neural networks," in *2020 42nd annual international conference of the IEEE engineering in medicine \& biology society (EMBC)*, 2020, pp. 4269–4272.

[26] J.-M. Valin, F. Michaud, and J. Rouat, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering," *Rob. Auton. Syst.*, vol. 55, no. 3, pp. 216–228, 2007.

[27] Q. Zhang *et al.*, "Cuff-less blood pressure measurement using pulse arrival time and a Kalman filter," *J. Micromechanics Microengineering*, vol. 27, no. 2, p. 24002, 2017.

[28] S. R. Chhatpar and M. S. Branicky, "Particle filtering for localization in robotic assemblies with position uncertainty," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 3610–3617.

[29] G. Li, X. Zeng, J. Lin, and X. Zhou, "Genetic particle filtering for denoising of ECG corrupted by muscle artifacts," in *2012 8th International Conference on Natural Computation*, 2012, pp. 562–565.

[30] A. Akbari, J. Martinez, and R. Jafari, "Facilitating Human Activity Data Annotation via Context-Aware Change Detection on Smartwatches," *ACM Trans. Embed. Comput. Syst.*, vol. 20, no. 2, Jan. 2021.

[31] J. Martinez, K. Sel, B. J. Mortazavi, and R. Jafari, "Boosted-SpringDTW for Comprehensive Feature Extraction of Physiological Signals," *IEEE Open J. Eng. Med. Biol.*, 2022.

[32] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. De Sousa, "Ensemble approaches for regression: A survey," *Acm Comput. Surv.*, vol. 45, no. 1, pp. 1–40, 2012.

[33] V. Nathan and R. Jafari, "Particle filtering and sensor fusion for robust heart rate monitoring using wearable sensors," *IEEE J. Biomed. Heal. informatics*, vol. 22, no. 6, pp. 1834–1846, 2017.

[34] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proc. IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[35] F.-F. Wei *et al.*, "Clinical significance of mean and pulse pressure in patients with heart failure with preserved ejection fraction," *Hypertension*, vol. 79, no. 1, pp. 241–250, 2022.

[36] G. S. Stergiou *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.

[37] N.-M. Z. Bakkar, A. F. El-Yazbi, F. A. Zouein, and S. A. Fares, "Beat-to-beat blood pressure variability: an early predictor of disease and cardiovascular risk," *J. Hypertens.*, vol. 39, no. 5, pp. 830–845, 2021.