

Data-Driven Guided Attention for Analysis of Physiological Waveforms with Deep Learning

Jonathan Martinez, *Student Member, IEEE*, Zhale Nowroozilarki, *Student Member, IEEE*,
Roobeh Jafari, *Senior Member, IEEE*, and Bobak J. Mortazavi, *Senior Member, IEEE*

Abstract— Estimating physiological parameters - such as blood pressure (BP) - from raw sensor data captured by noninvasive, wearable devices rely on either burdensome manual feature extraction designed by domain experts to identify key waveform characteristics and phases, or deep learning (DL) models that require extensive data collection. We propose the Data-Driven Guided Attention (DDGA) framework to optimize DL models to learn features supported by the underlying physiology and physics of the captured waveforms, with minimal expert annotation. With only a single template waveform cardiac cycle and its labelled fiducial points, we leverage dynamic time warping (DTW) to annotate all other training samples. DL models are trained to first identify them before estimating BP to inform them which regions of the input represent key phases of the cardiac cycle, yet we still grant the flexibility for DL to determine the optimal feature set from them. In this study, we evaluate DDGA's improvements to a BP estimation task for three prominent DL-based architectures with two datasets: 1) the MIMIC-III waveform dataset with ample training data and 2) a bio-impedance (Bio-Z) dataset with less than abundant training data. Experiments show that DDGA improves personalized BP estimation models by an average 8.14% in root mean square error (RMSE) when there is an imbalanced distribution of target values in a training set and improves model generalizability by an average 4.92% in RMSE when testing estimation of BP value ranges not previously seen in training.

Index Terms—Blood pressure, deep learning, dynamic time warping, guided attention

I. Introduction

Continuous monitoring of physiological parameters throughout the varying contexts of patients' daily lives enables the diagnosis of life-threatening illnesses in their early stages when they may not have been made through infrequent clinical measurements [1]–[5]. Noninvasive, wearable devices, such as smartwatches or smartrings, yield comfortable and continuous collection of physiological waveforms from which key health parameters are derived. Yet, estimation of some parameters – such as blood pressure (BP) - require non-linear modeling from the captured modalities due to evolving

morphological characteristics associated with personalization [6], aging [7], and overall health status [8]. Therefore, domain experts often develop task-specific estimation algorithms that depend on burdensome comprehensive feature selection from manually annotated waveform fiducial points – morphological characteristics that represent key physiological events [9]–[13]. Yet, gaining a deep understanding of how novel wearable modalities describe human physiology requires laborious experimentation to ultimately determine the morphological characteristics from which feature sets should be extracted. Deep learning (DL) models alleviate these challenges with automatic feature extraction, although, optimization requires extensive data collection of a wide range of target values [14], which is impractical in the remote health-science domain. Thus, we propose a BP estimation framework leveraging minimal expert-knowledge (minimal expert annotation) to guide DL models' training process ensuring domain-aware automatic feature extraction. Furthermore, our approach balances the fusion of task-specific and DL-based approaches to achieve personalized BP estimation, and enhance model accuracy and generalizability.

Task-specific estimation models depend on expert understanding of the wearable waveforms to identify and derive the optimal set of features that can be mapped to the target physiological parameter. For example, with the BP estimation task, health-scientists have identified pulse transit time (PTT) or pulse arrival time (PAT) as a critical feature that captures the rate of blood flow throughout the body [15], [16]. In addition, a number of supplementary temporal and morphological features have been proposed and can be derived from fiducial points that represent key phases of the cardiac cycle (diastolic pressures, systolic pressures, and various reflections on arterial compliance) [17] – such as heart rate, interbeat interval, rise time, fall time, area under the curve, amplitude, etc. All of which, are input to various types of non-linear regression models for BP estimation [18]. Yet, the viability of specific feature sets does not necessarily generalize to all types of estimation models neither across patients. Such feature extraction may also lead to loss of information if the feature set

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)

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

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

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

is not large enough to adequately describe the whole waveform.

Alternatively, DL models have been proposed to achieve automatic feature extraction by iteratively analyzing a training dataset of sample waveforms without any required additional burden on domain experts to extract supplemental features. Particularly, architectures based on convolutional neural network (CNN) [19], recurrent neural network (RNN) [20], and transformer modules [21] have displayed a strong ability to capture spatial and temporal patterns of physiological waveforms that may not necessarily be inherently captured by the feature sets previously proposed and extracted by hand-crafted algorithms [22]. In addition, attention mechanisms have shown significant improvements to this task by focusing model analysis only on the regions of the input which contribute the most to estimation of physiological parameters like BP [23]. However, if the size and variance of the training data set is not sufficient, deep learning-based architectures may learn secondary features that are overfit to target value samples that compose the majority of those seen in training, i.e. the mean of training samples (120/80 mmHg for BP). Thus, models may not generalize to predicting target values that exist in the minority region or were not previously seen in training. This phenomenon of the DL-domain is commonly referred to as bias in the training dataset toward the majority regions [24].

Guided attention is a proposed solution that incorporates domain expertise into the DL training process to impose onto the models which features of the input should be used for prediction [25]. In the computer vision domain, this is often achieved by optimizing the learned attention maps that reflect which features of the input to the model has been identified to contribute the most to the output prediction [26]. In the context of an object classification task, this solution will encourage the DL model to focus most attention on the object to be classified as opposed to any background features present in the image. However, this type of solution does not directly translate to time-series data as model attention should not be focused onto one sample of an input waveform but rather will often need to consider the relationship across all points of the input. Furthermore, it has not yet been well established or studied how this knowledge is reflected into attention maps. Physics-guided neural networks are an alternative approach that introduces sub-tasks for the DL associated with feature engineering. But, these approaches impose the models to learn specific feature sets that depend on domain-expertise as opposed to granting it total flexibility to discover optimal feature sets, thus limiting model capabilities [25], [27].

In this work, we propose a novel Data-Driven Guided Attention (DDGA) framework for optimally guiding DL models in training with minimal expert annotation – requiring the manual annotation of only a single (optimal) cardiac cycle's associated waveform. We achieve a novel integration of proven signal processing techniques to enable researchers to communicate information to the DL-based estimation models to enhance performance through a novel Guided Attention solution that is best fit for time-series data. We train DL feature extraction and model attention components to be aware of key fiducial points in an input physiological waveform that are

meaningful to BP estimation tasks as determined by domain expertise, however, we do not restrict them to learn specific feature sets from them. We grant the model enough flexibility to consider the whole input waveform to determine optimal features. Our method achieves a balance between domain-specific feature extraction and automated learning while leveraging the corresponding benefits of each to yield enhanced personalized BP estimation and enhanced personalized model generalizability when testing over new ranges of BP out-of-distribution from the training set target values. In testing, BP estimation is achieved with the optimized DL architecture that requires only the raw physiological waveforms collected by wearables. The contributions of this work are as follows:

- We propose a novel DDGA integration that incorporates minimal domain expertise into the DL model training process to improve the feature sets automatically extracted by them.
- We introduce a Guided Attention solution that is best fit for time-series data types, where limited exploration has been conducted.
- We demonstrate the proposed DDGA framework's ability to enhance personalized DL-based BP estimation models and their generalizability to out-of-distribution target values through experimentation with two datasets of distinct physiological modalities.

II. RELATED WORK

The most prevalent DL-based architectures proposed for analysis of physiological waveforms for BP estimation - such as electrocardiogram (ECG), photoplethysmography (PPG), and ballistocardiogram (BCG) - include bi-directional long short-term (Bi-LSTM) neural networks [20], a combination of CNN and Bi-LSTM neural network layers with self-attention [23], and transformer (attention-based) layers [21], [28]. However, most previous work that specifically target BP estimation tasks incorporate supplementary pre-processing or feature selection based on domain-knowledge to guide model optimization.

Previously proposed BP estimation frameworks have implemented hand-crafted algorithms to first segment ECG and PPG waveforms into cardiac cycles before extracting several waveform features that were ultimately analyzed by Bi-LSTM layers to capture their temporal information – referred to as multi-stage analysis [28]–[30]. Thus, heavily relying on domain-expertise and not leveraging the full benefit of DL-based models. Similarly, a follow up work adapted a similar pre-processing step to segment PPG waveforms into cardiac cycles, however, leverage automatic feature extraction by CNN layers to the segmented waveforms as opposed to hand-crafted algorithms before analysis by Bi-LSTM layers [31]. Alternative BP estimation solutions do not require the pre-segmentation step for analysis and allow DL models to directly analyze time-based windows of raw waveforms, yet jointly input supplementary features in an attempt to benefit from both types of approaches. Particularly, one type of approach inputs the

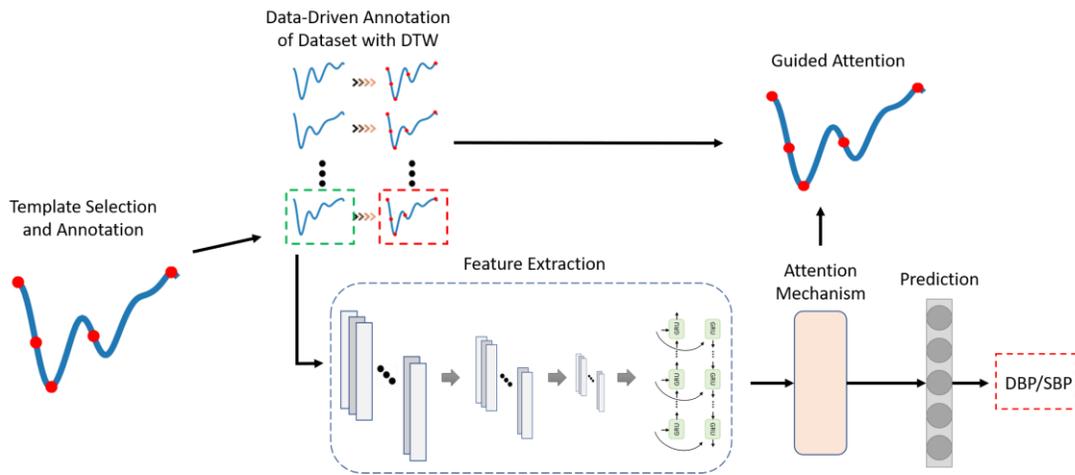


Fig. 1. End-to-end view of the Data-Driven Guided Attention Framework (DDGA) for personalized DBP/SBP estimation.

same morphological time- and frequency-domain features that earlier work have proposed in an attempt to guide DL of relevant regions of the raw ECG, PPG, and BCG waveforms [18]. Other approaches rather jointly input patient physical characteristics and other biometrics in an attempt to help estimation models generalize [19], [32]. However, all above approaches are still heavily dependent on expert knowledge to design robust feature extraction algorithms (which is not always feasible for novel wearable waveforms) and to identify the appropriate set of patient information to achieve high-quality BP estimation. More recent studies have leveraged multi-task driven approaches by training models to estimate both SBP/DBP simultaneously in an attempt to automatically guide feature extractors to identify task-specific and task-agnostic embeddings without any hand-crafted pre-processing [22], [33]. However, they then fall short of leveraging any domain expert guidance thus still leaving them susceptible to the aforementioned challenges associated with learning-based models. In this study, we show that our proposed DDGA framework can leverage the benefits of DL’s automated feature extraction and guidance from minimal domain expertise, while analyzing only raw waveforms and without the need for additional hand-crafted pre-processing steps or the addition of supplementary features manually selected by domain experts.

III. Methodology

In Fig. 1, we show an end-to-end illustration of our proposed DDGA framework. First, only a single template waveform is manually selected by the researcher to represent the general case of the input physiological waveform morphology. Key fiducial points of the waveform that represent key phases of the cardiac cycle are annotated onto this template waveform only, where dynamic time warping (DTW) is then leveraged to compare the template to all other samples in the training dataset, identifying all fiducial points in each automatically without additional manual annotation. Then, Guided Attention is employed during estimation model optimization, where baseline DL architectures are taught to first identify where these key fiducial points exist in the input waveform before estimating BP – using the DTW annotations as labels. Finally, in testing the BP estimation model receives only the

physiological waveforms collected by wearables as input, and is capable of identifying the key fiducial points without any annotated assistance before extracting the learned feature sets from training to map them to a strong BP estimation. In what follows, we describe the process of building this framework and achieving Guided Attention during the training process: annotating the signals (DTW) and training DL-based estimation models through Guided Attention. Our primary innovation of this work is the novel integration of strong independent signal processing techniques in such a way that enables researchers to communicate information into the DL architectures during training. We apply our model to two case studies (See Section IV): 1) on the MIMIC-III waveform dataset [34] to serve as a benchmark for performance under the standard BP estimation experimental settings with 5-sec windows of PPG as defined in previous work [18], [23]; and 2) on a bio-impedance (Bio-Z) dataset [17] under the beat-to-beat setting, which also contains less training data per patient to demonstrate DDGA’s improved optimization of DL models under this limitation.

A. Data-Driven Annotation with DTW

Annotating the key fiducial points for the Bio-Z and PPG waveforms analyzed was achieved with Dynamic Time Warping (DTW). This first involves the manual selection and annotation of a single template waveform that represents a single cardiac cycle and will be compared to all other samples for all other subjects/patients in the dataset. In Fig. 2 we show the two selected templates used in this study for a single PPG and a Bio-Z cardiac cycle with annotated onset, systolic, diastolic, max slope, and inflection fiducial points that reflect key phases of the cardiac cycle [17], [35]. For this study, template waveforms were selected manually by a domain expert as instances that adequately represent the general morphology of each modality with relatively minimal noise. This task requires only minimal understanding of how cardiac cycles are represented in the generated physiological modality, and only takes place once as the template may be reused for all subsequent patients’ data. Through this automated fiducial point annotation approach, our solution is able to leverage the benefits of domain expertise while almost removing all burden associated with hand-crafted design of feature extraction algorithms. In this work, the domain expert randomly

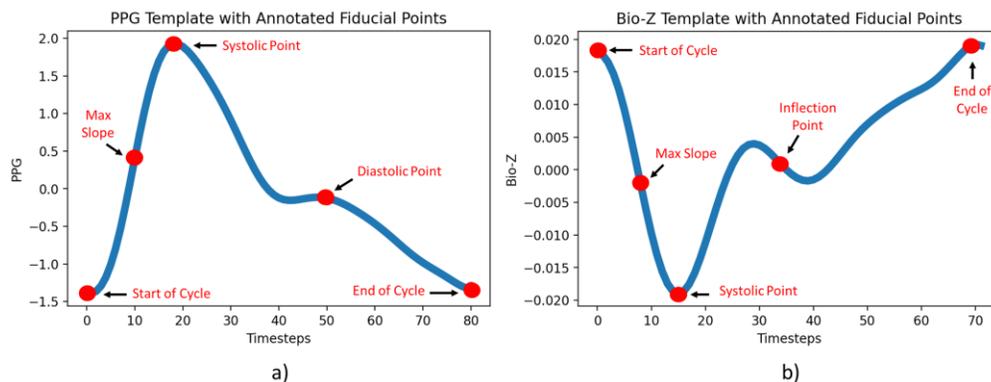


Fig. 2. Cardiac cycle template waveforms with annotated fiducial points for a) PPG from MIMIC-III and b) Bio-Z.

subsampled a manageable number of waveform instances to be reviewed for template selection. Alternatively, previous work leveraging DTW comparisons have generated template waveforms through ensemble averaging of multiple instances [36] present in a time window. However, since annotation is our primary goal and depends on precise sample-to-sample mappings, we chose the former approach to avoid any waveform corruption which could be associated with ensemble averaging. Furthermore, previous work that have evaluated DTW's annotation strength have demonstrated this type of approach to be robust to evolving waveform morphologies due to both intra- and inter-subject variation [37]. The fiducial points to be annotated were determined based on those which are typically used for feature extraction in previous work that propose domain-specific BP estimation algorithms [17], [38], [39].

DTW yields the optimal sample-to-sample mappings between two waveform instances (Bio-Z or PPG). Given the designated template waveform, $V = v_1, v_2, \dots, v_M$ where M is its length, whose annotated fiducial point locations are tracked in F_V , and given an input waveform to be analyzed, $X = x_1, x_2, \dots, x_T$ where T is its length, a distance matrix is constructed through a series of comparisons [40]

$$d(x_t, v_m) = \|x_t - v_m\|_2^2 + \min \begin{cases} d(t, m - 1) \\ d(t - 1, m) \\ d(t - 1, m - 1) \end{cases} \quad (1)$$

$$\forall t \in (1, T), \forall m \in (1, M)$$

Where $d(x_t, v_m)$ describes the DTW comparison between the t^{th} sample of the input waveform and the m^{th} sample of the template. The distance is computed as the Euclidean distance, $\|x_t - v_m\|_2^2$, added to the smallest adjacent comparison in the distance matrix. After all comparisons have been made, a warping path, $W = w_1, w_2, \dots, w_K$ where K is its length, is back traced to obtain the optimal alignment between the two waveforms where a given w_k represents a sample-to-sample mapping. Starting from the last distance computation, we will define a shortest path towards the first by indexing the minimum adjacent values – $d(t, m - 1)$, $d(t - 1, m)$, and $d(t - 1, m - 1)$. Last, annotation is ultimately achieved by identifying the points in X that map to the fiducial points, $F_V \subseteq X$, according to the warping path. In Fig. 3 we visualize this – using Bio-Z as an example in this case although the approach is the same for the PPG waveforms. After the DTW distance matrix is constructed with all pairwise distances (grey), we can back trace it to obtain the warping path (red) which maps each waveform step to its optimal counterpart (such as visualized by the yellow point). Thus, by only requiring a single template

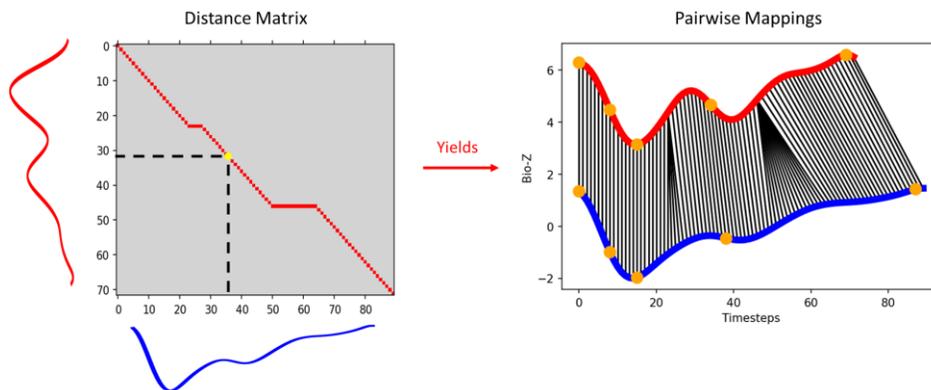


Fig. 3. a) Example DTW Distance Matrix with the mapped warping path (red) which yields the b) sample-to-sample optimal mappings between two waveforms. A given distance value (yellow) links the timesteps for each waveform as optimal counterparts.

cardiac cycle to be annotated by a domain expert, this approach completely automates the annotation of the remaining waveform instances.

The Bio-Z dataset was already segmented into its cardiac cycle waveforms when it was curated, therefore, the annotation implementation is straightforward as formulated above and we leverage the FastDTW [41] which optimizes standard DTW's runtime. However, for the PPG waveforms of the MIMIC-III dataset, since each instance is a 5-second window of waveform potentially representing multiple cardiac cycles, we leverage the Boosted-SpringDTW [37] framework which adapts DTW to first segment quasi-periodic physiological signals into cardiac cycles before then extracting the sample-to-sample comparisons. Particularly, this framework assigns a probability to each sample in the 5-second stream to describe the likelihood that each is a cardiac cycle endpoint (as formulated in the original work)

$$P(e_t) = P(c_t|x_t) * P(d(x_t, v_m)) \quad (2)$$

$$P(c_t|x_t) = \frac{g_t - \min(X')}{\max(X') - \min(X')} \quad (3)$$

$$P(d(x_t, y_m)) = e^{-\gamma * d(x_t, y_m)} \quad (4)$$

Where c_t are candidate endpoints of X if they are local minima points in the waveform, and they are scored according to the steepness of the immediately following max slope point detected in the first derivative of the physiological waveform, g_t . Determining candidate endpoints is based on the basic understanding of how PPG is generated to reflect blood flow, therefore, $P(c_t|x_t)$ can be considered the likelihood based on morphological characteristics of a given waveform. Then, $P(d(x_t, y_m))$ is the likelihood that leverages DTW. Particularly, $d(x_t, y_m)$ is the last distance computed at a given time step of the given waveform stream, and this reflects its current overall similarity to the template. This value should be minimized when a completed cardiac cycle has been encountered, therefore we progressively reward the distances with a monotonically increasing function such as the shown exponential function. Last, we identify the local maxima likelihood pairs that would define a detected cardiac cycle,

$P(e_t)$, while respecting the plausibility constraint that its expected length corresponds to the estimated local heart rate as obtained by the waveforms frequency components. Last, once segmentation is complete, annotation may be achieved as previously discussed using the distance matrix's warping path.

Further details regarding this method are found in the cited work which proposed Boosted-SpringDTW. This citation also includes sufficient experimentation which validates DTW's viability for annotation tasks, and also demonstrates that high-quality annotation may be achieved with single template initialization. For both Bio-Z and PPG, we used the first derivative of waveforms for comparison and normalized each instance to zero mean and unit variance.

B. Guided Attention with Multi-Task Learning

Guided Attention takes place only during the model training process to optimize its feature extraction and model attention components. In Fig. 4, we show a conceptual model of the proposed Guided Attention learning framework. We organized components of a DL architecture into that which is typically adapted by state-of-the-art BP estimation models a) Feature Extraction, b) Attention Mechanism, and c) Prediction, and later demonstrate improvement over several common architecture types (discussed more in the experiments and results section). The Feature Extraction component of the framework received only the waveforms as input to encode various waveform and time characteristics to a hidden state, and its output was then analyzed by the Attention Mechanism to identify critical regions and model dependencies across timesteps of the input. In traditional models, this output was then directly passed to the Prediction component of the model to map to the estimated BP values. However, for Guided Attention learning, the output of the Attention Mechanism was simultaneously fed to a separate output layer to predict the locations of fiducial points in the waveform inputs. That is, it was decoded to a binary vector of the same length as the waveform input, where a 1 indicates the location of a key fiducial point and a 0 indicates the location of a non-fiducial point. Here, the fiducial point annotations extracted with DTW are considered as the ground truth labels for this sub-task. Both tasks were optimized simultaneously in training with a joint

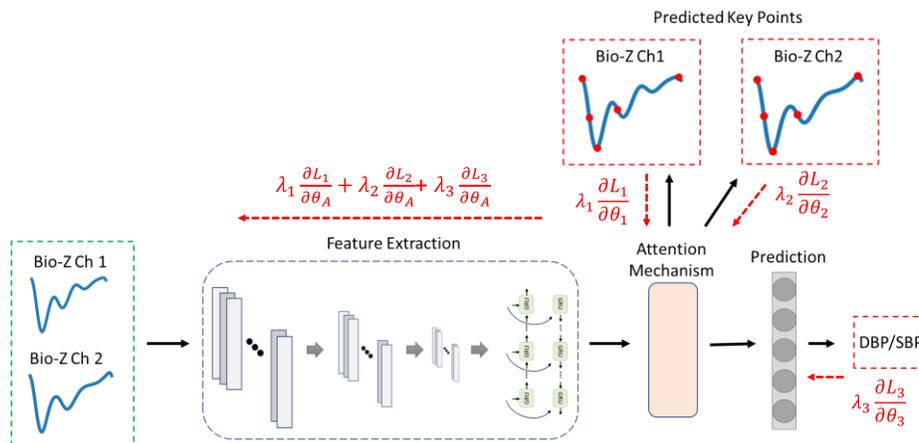


Fig. 4. Conceptual model showing the Feature Extraction, Attention Mechanism, and Prediction phases of a standard framework. DDGA is applied to the output of the Attention Mechanism to ensure that the modelled dependencies include those related to the key fiducial points.

objective function. This loss function was a combination of binary cross-entropy (BCE) for training the fiducial points, and mean squared error (MSE) for BP estimation. We chose MSE for regression optimization as opposed to other metrics since it is the most punitive loss function that would more rapidly encourage the model to balance all tasks. The joint objective function was then composed as

$$L_{joint} = \lambda_1 L_1 + \lambda_2 L_2 + \lambda_3 L_3 \quad (5)$$

where λ_1 , λ_2 , and λ_3 were multipliers assigned to each sub-task (1 for channel 1, 2 for channel 2 and 3 for BP estimation) in the joint loss function to convey to the DL model the notion of task importance.

Guided Attention enables knowledge injection into the DL through the backpropagation process where model parameters are updated. First, each of the Prediction layers are updated as

$$\theta_1 = \theta_1 + \alpha \cdot \lambda_1 \frac{\partial L_1}{\partial \theta_1} \quad (6)$$

$$\theta_2 = \theta_2 + \alpha \cdot \lambda_2 \frac{\partial L_2}{\partial \theta_2} \quad (7)$$

$$\theta_3 = \theta_3 + \alpha \cdot \lambda_3 \frac{\partial L_3}{\partial \theta_3} \quad (8)$$

where θ_1 , θ_2 , and θ_3 represent the Prediction layer parameters associated with each task, and α is the learning rate. Then, the gradients of each are summed to each provide influence onto the Attention Mechanism as

$$\theta_A = \theta_A + \alpha \cdot \left(-\lambda_1 \frac{\partial L_1}{\partial \theta_A} + \lambda_2 \frac{\partial L_2}{\partial \theta_A} + \lambda_3 \frac{\partial L_3}{\partial \theta_A} \right) \quad (9)$$

where θ_A represents the Attention Mechanism parameters. Thus, all knowledge associated with each task is transferred through the computed gradients, and the Attention Mechanism will jointly consider all simultaneously while prioritizing tasks as imposed by the importance multipliers (λ_1 , λ_2 , and λ_3). This formulation can support any additional or alternative loss functions that pursue optimization of classification and regression tasks. The addition or exchange for other sub-tasks would only require additional tuning of the associated multipliers to manage task importance.

In this work, we estimate diastolic and systolic blood pressure (DBP and SBP) independently which is a common practice with traditional techniques due to distinct correlations amongst feature sets and DBP/SBP [42]. Furthermore, isolating DBP and SBP to separate estimation models further enhances the impact of Guided Attention during optimization since it alleviates the need to balance task importance for each sub-task by the model. The multipliers may be tuned to maximize model performance by ensuring that a balance is achieved to consider the key fiducial point identification sub-task as critical enough to necessitate high-quality predictions while still understanding that the main objective with greatest priority is the BP estimation task. Achieving this balance was especially important as the order of magnitude for L_1 , L_2 , and L_3 for BCE and MSE are very different and are dynamic throughout training. This approach ensures that the set of features

prioritized by the Attention Model will include those derived from the fiducial points.

IV. EXPERIMENTS AND RESULTS

We evaluated the impact of DDGA on three of the most prominent cuffless BP estimation and DL-based regression baseline architectures when analyzing two cuffless BP datasets: MIMIC-III waveform benchmark dataset and the Bio-Z dataset. We include MIMIC-III since it includes the standard conditions adapted in previous work and it is a benchmark dataset with ample training data. The input instances are 5-second windows of single-channel PPG data where each model aims to estimate the average BP value. In the contrary, the Bio-Z dataset contains novel physiological waveforms collected with an innovative wrist wearable technology, and it also possesses the conditions by which there is not necessarily an abundance of training data per subject. Each input instance contains two channels of Bio-Z waveforms that represent a single cardiac cycle, where the aim is to estimate the corresponding BP value for it. We evaluate each baseline architecture before and after applying DDGA to estimate DBP/SBP independently for each dataset under two conditions: 1) traditional DL training using all ranges of BP available per subject, and 2) interpolation and extrapolation experiments where BP ranges were iteratively removed from each subjects' training data to be used solely for testing. Thus, we are able to evaluate DDGA's impact on the standard personalized BP estimation task and also on its ability to generalize to new BP ranges without calibration.

Root mean squared error (RMSE) was used to evaluate the precision of each model since it is measured in the same units as BP – millimeters of mercury (mmHg). We also include the Pearson correlation coefficient to ensure that model predictions are increasing/decreasing accordingly with the various ranges of BP. This better reflects if a model is only learning to estimate the average values of the BP range observed in the training data, which is reflected with a lower correlation score. While these metrics were evaluated using all of the predictions and test data for each subject per each experimental setting, we also include the average change in RMSE and correlation per subject to measure the amount of impact DDGA yields per each personalized model.

A. Cohort Selection for MIMIC-III

The complete MIMIC-III waveform database consists of over 2.4k patients' data while admitted into the intensive care unit (ICU) of varying demographics and varying degrees of health states. We defined 4 criteria to select a cohort of patients from this total set, based on 1) age, 2) health condition, 3) length of stay in the ICU, and 4) waveform quality. We selected patients between the ages of 18 to 65 in an effort to capture adults whose collected PPG waveforms maintain a strong reflection of cardiac activity. The age of 65 was determined as the upper bound since this is the age which chronic heart conditions begin to develop [43]. Next, we selected patients who were not suffering from a cardiovascular condition (such as myocardial infarction, hypertension, heart arrhythmia, etc.), major internal hemorrhaging, or vital organ failure, to ensure that a strong relationship between the collected PPG waveforms and BP exist. Next, we selected patients whose ICU stay did not

exceed 3 days since it has been shown in previous research that mortality rates increase to 30% after a 3-day ICU stay and will continue to rise accordingly [44]. Last, we applied waveform quality constraints to these queried patients' data [18]. They include a minimum stream length of 10 minutes to adequately capture some level of BP variation, flat line and flat peak removal, and waveform outlier removal based on a Hampel filter.

Applying the above criteria yields 31 patients of this dataset to be included in the evaluation of these experiments. There were 13,636 valid cardiac cycles on average per patient where the maximum number of available instances for a patient was 32,689, the minimum number of available instances for a patient was 3,068, and the standard deviation was 7,907. Therefore, this benchmark dataset provides a wide range of data availability conditions thus providing a thorough evaluation of each of the models to be compared in our experiments.

B. Description of Bio-Z System

The dataset used in this study is comprised of Bio-Z signals collected from 11 healthy participants under the IRB approval by IRB2017-0086D at Texas A&M University [17]. Bio-Z is an electrical signal which can be measured by injecting a current into the human body and capturing the voltage difference across the radial artery, thus yielding a two-channel waveform. The variation in Bio-Z corresponds to changes in blood pressure over time at the position where the sensor is placed. In order to capture a wider range of blood pressure during collection, subjects were asked to conduct physical activities during successive trials. After discarding regions of data collection that were corrupted due to noise or lack of sufficient contact of the sensor with the skin, there were 1,057 valid cardiac cycles on average per subject where the maximum number of available instances for a subject was 1,917, the minimum number of available instances for a subject was 440, and the standard deviation was 437. As observed, the amount of available training instances for this dataset is significantly less than that of the MIMIC-III benchmark dataset. Therefore, this dataset allows us to evaluate DDGA performance in the presence of two unique challenges: 1) when analyzing a novel physiological modality whose available domain expertise is relatively less than that of PPG, and 2) when the amount of training instances used for model optimization is very small. Both of which could potentially lead standard DL-based estimation models to underperform and fail to generalize.

C. Baseline Architectures Tested

We evaluated each model's ability to estimate BP directly from input waveforms before and after applying our proposed DDGA on commonly-used deep learning architectures. We evaluate only neural network-based architectures that pursue automated feature extraction since they inherently do not consider domain knowledge. Applying DDGA to other types of machine learning solutions such as tree-based or SVM models would not be practical since feature extraction algorithms designed for these cases typically involve domain expert intervention for feature extraction, therefore DDGA optimization would not be necessary in these cases. Although, for comparison, we did include a multi-layer perceptron (EF-MLP) baseline approach for each modality (PPG and Bio-Z)

that uses feature sets obtained by extraction algorithms hand-crafted by domain experts [10], [12], [17], [19], [28], [45]. The feature sets included time-, amplitude-, and area-based characteristics to describe the waveform morphologies (12 Bio-Z and 11 MIMIC-III for each cardiac cycle). The MLP consisted of three dense layers with 2000, 1000, and 1 hidden units respectively. The automated feature extraction models for this experiment include: 1) Bi-directional Gated Recurrent Unit with Self-Attention (Bi-GRU+Attn) [23], 2) a hybrid 1-Dimensional Convolutional Neural Network and Bi-directional Gated Recurrent Unit with Self-Attention (CNN+Bi-GRU+Attn) [23], and 3) the Transformer [21] model. The first model inputs the raw waveforms into a single Bi-GRU layer with 64 hidden units, whose complete hidden matrix output corresponding to all time steps is received by a multiplicative self-attention layer and ultimately flattened and passed through a single output dense layer that maps the extracted features to a single estimated BP value. The second model includes the same architecture from the former but precedes it with three CNN modules composed of a single CNN layer (where the number of feature maps increases for each – 32, 64, and 128), a batch normalization layer whose output undergoes rectified linear unit activation, and a max pooling layer with a factor of 3. The third model first applies positional encoding of the raw waveforms before both are input into a 4-headed self-attention block consisting of the dot-product attention mechanism with head sizes of 64 units followed by layer normalization, where the output is flattened and passed through the single output dense layer which maps extracted features to the estimated BP value.

Estimation performance was evaluated for each of the architectures before and after applying DDGA. The same networks were used for each BP estimation scenario (i.e., personalized, interpolation and extrapolation). Each model was trained for 50 epochs where validation error was tracked to store the weights of the epoch that yielded the lowest validation error – to then be used for testing. The standard implementation of the Adam optimizer was used with a fixed learning rate of 0.001 for all architectures to ensure fair comparison, however, batch sizes were set as 128 for MIMIC-III and 16 for Bio-Z to adapt to the differing sizes in the number of training instances. The task importance multipliers (λ_1 , λ_2 , and λ_3) were obtained through a grid search optimization where a pre-defined range of values, [0.05, 0.25, 0.5, 0.75, 1.0], were iteratively evaluated and the best performing combination was selected. However, the λ_1 and λ_2 multipliers which are applied to each channel of the same physiological modality (such as for Bio-Z) were set to the same value since fiducial point identification for each may be considered as one task together.

D. Improved Personalized BP Estimation with DDGA

The personalized models were individually evaluated using only a given subject's data for training and testing, which is consistent with the formulation by previous work and has been shown to be the most effective approach for non-invasive BP estimation with wearables [17], [18], [22], [46]. For each, 80% of a subject's data was set aside for training (10% of which was held out for validation), and the remaining 20% of data was used for testing. Therefore, there were no redundant samples present in both training and testing. These training and test

TABLE I
PERSONALIZED DBP/SBP ESTIMATION PERFORMANCE FOR BASELINE ARCHITECTURES WITH DDGA

Dataset	Model	DBP			SBP		
		RMSE (mmHg)	$\overline{\Delta RMSE}$ (mmHg)	r	RMSE (mmHg)	$\overline{\Delta RMSE}$ (mmHg)	r
MIMIC-III	EF-MLP	6.32 ± 9.18	-	0.79	11.33 ± 15.75	-	0.82
	Bi-GRU+Attn	5.82 ± 8.34	-	0.82	8.94 ± 12.74	-	0.89
	DDGA+Bi-GRU+Attn	5.73 ± 8.20	-0.13	0.83	8.68 ± 12.52	-0.27	0.90
	CNN+Bi-GRU+Attn	5.47 ± 8.07	-	0.85	8.69 ± 12.74	-	0.90
	DDGA+CNN+Bi-GRU+Attn	5.34 ± 7.89	-0.14	0.85	8.48 ± 12.14	-0.22	0.90
	Transformer	6.97 ± 9.73	-	0.73	12.14 ± 16.42	-	0.78
Bio-Z	DDGA+Transformer	6.89 ± 9.64	-0.05	0.74	12.07 ± 16.39	-0.08	0.79
	EF-MLP	5.21 ± 6.97	-	0.80	7.49 ± 9.93	-	0.78
	Bi-GRU+Attn	5.61 ± 7.42	-	0.72	7.80 ± 10.24	-	0.77
	DDGA+Bi-GRU+Attn	4.91 ± 6.19	-0.63	0.79	7.19 ± 9.06	-0.58	0.80
	CNN+Bi-GRU+Attn	4.93 ± 6.22	-	0.78	6.87 ± 8.73	-	0.82
	DDGA+CNN+Bi-GRU+Attn	4.32 ± 5.36	-0.54	0.84	6.63 ± 8.35	-0.24	0.84
Transformer	5.19 ± 6.53	-	0.76	7.06 ± 8.97	-	0.81	
DDGA+Transformer	4.65 ± 5.79	-0.58	0.82	6.90 ± 8.67	-0.17	0.82	

splits were used for both the implementation with and without DDGA for fair comparison. Furthermore, the Bio-Z data samples were shuffled before train/test splitting since all extracted input-output instances were standalone cardiac cycles – as there was no notion of time associated with each cardiac cycle. Therefore, we handle the MIMIC-III data samples in the same manner for consistency. It is only over the designated training data which waveform annotation and further Guided Attention were applied to optimize the DL architectures. In testing, only the optimized DL model is used to estimate BP – which receives only the physiological waveforms collected by wearables as input.

In Table 1, we show DBP and SBP performance for each baseline architecture and for the improvement observed with DDGA for both datasets. First, we measure performance with respect to RMSE (and their 95% confidence intervals) and Pearson’s correlation coefficient (r), where we calculate by combining all test data together for all subjects. Then, we provide the average improvement to RMSE for each subject after applying DDGA to each baseline model. Thus, for the average improvement for RMSE, the more negative result is the superior performer. Although marginal, we observe improvements to each baseline architecture for the MIMIC-III

dataset for both DBP and SBP up to 2.91%. For the Bio-Z dataset, we observe a much larger impact by DDGA compared to the MIMIC-III experiments. This is the case since the MIMIC-III dataset contains a much larger amount of training instances compared to the Bio-Z dataset. Furthermore, the ranges of BP values in the former are wider since the patient cohort may have contained ill health conditions. Therefore, DDGA observes further improvement with the Bio-Z dataset since the amount and variance of its training instances is more limited – which is the more practical scenario for future remote health monitoring tasks. Thus, DDGA helps the estimation models generalize. We observe improvements to all baseline architectures up to 12.5%, also achieving estimation performance competitive and in some cases better than those obtained with the benchmark dataset. Furthermore, we observe approximately twice the magnitude of BP estimation improvement for the average improvements per subject. Compared to the baseline approach that leverages extracted features, almost all DDGA architectures outperformed the baseline, thus confirming that leveraging both domain expertise and automated feature extraction yields higher precision.

In Tables II and III we also show the fiducial point identification quality by each of the DL-based models

TABLE II
ANALYSIS ON FIDUCIAL POINT IDENTIFICATION PERFORMANCE FOR MIMIC-III

Dataset	Model	DBP	SBP
		\overline{BCE}	\overline{BCE}
MIMIC-III	DDGA+Bi-GRU+Attn	0.706	0.706
	DDGA+CNN+Bi-GRU+Attn	0.706	0.706
	DDGA+Transformer	0.706	0.706

TABLE III
ANALYSIS ON FIDUCIAL POINT IDENTIFICATION PERFORMANCE FOR BIO-Z

Dataset	Model	DBP		SBP	
		\overline{BCE}_1	\overline{BCE}_2	\overline{BCE}_1	\overline{BCE}_2
Bio-Z	DDGA+Bi-GRU+Attn	0.706	0.707	0.706	0.707
	DDGA+CNN+Bi-GRU+Attn	0.707	0.708	0.707	0.708
	DDGA+Transformer	0.707	0.707	0.706	0.706

TABLE IV
ANALYSIS ON PERSONALIZED DBP ESTIMATION PERFORMANCE THROUGHOUT DISTRIBUTION QUARTILES

Dataset	Model	DBP				SBP			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
		$\overline{\Delta RMSE}$							
MIMIC-III	DDGA+Bi-GRU+Attn	0.02	-0.15	-0.48	-0.83	-0.44	-0.01	-0.61	-0.98
	DDGA+CNN+Bi-GRU+Attn	-0.16	-0.07	-0.19	-0.48	-0.19	-0.20	-0.47	-0.20
	DDGA+Transformer	0.07	-0.04	-0.16	-0.32	-0.16	0.07	-0.10	-0.58
Bio-Z	DDGA+Transformer	-1.14	-0.56	-0.31	-1.27	-0.43	0.21	-0.96	-1.50
	DDGA+CNN+Bi-GRU+Attn	-0.55	0.20	-0.71	-1.15	0.01	-0.01	-0.24	-0.56
	DDGA+Transformer	-0.83	-0.43	-0.42	-1.43	-0.36	-0.06	-0.15	-0.28

according to the annotations generated by DTW. This task was evaluated through BCE which is the same metric leveraged during joint optimization. Despite that we have imposed a lower task priority for fiducial point identification compared to the BP estimation task, each model maintains robust performance and generalizes in testing. However, it should further be noted that this does not reflect the annotation quality by DTW as this would require manual annotation of all dataset instances to create a ground truth. Furthermore, the aim of this work is to leverage “data-driven” approaches, thus we rely on the proven effectiveness of DTW in previous work to evaluate the overall contribution to Guided Attention which has been demonstrated through the improved BP estimation task – therefore we refer to the previous work for demonstration of annotation quality amidst evolving waveform morphologies [37]).

E. Model Estimation in Various BP Regions

Aside from general trends in BP estimation, we additionally evaluated performance in various ranges of BP. This analysis better captures DDGA’s ability to help DL models generalize to ranges of target values that existed in the minority regions of the training dataset (in other words, when moving away from the common 120/80 mmHg value). For the MIMIC-III dataset, the distribution of DBP values have a range from 50.1 to 119.9 mmHg with a mean value of 65.9 mmHg and a standard deviation of 10.3, while the SBP value distribution has a range from 80.0 to 199.7 mmHg with a mean value of 124.8 mmHg and a standard deviation of 19.6. For the Bio-Z dataset, the distribution of DBP values have a range from 47.2 to 97.8

mmHg with a mean value of 72.0 mmHg and a standard deviation of 7.96, while the SBP value distribution has a range from 81.9 to 159.9 mmHg with a mean value of 120.7 mmHg and a standard deviation of 12.1. Thus, the majority of DBP/SBP values per patient existed around the mean. This can be harmful to DL-based models as it may cause them to overfit to this range of values by learning secondary features that do not necessarily allow its predictions to generalize.

Each patient’s test data was sorted and split by distribution quartiles – 4 quarters. In Table IV, we provide the average improvement (per patient) in RMSE which DDGA provides to each baseline model for each bucket range of DBP/SBP, respectively. In general, we observe an overall improvement with DDGA for most cases, particularly to the lower and upper ranges (First Quarter and Fourth Quarter) that were underrepresented in training.

F. Interpolation and Extrapolation

To further evaluate the generalizability of such personalized models we conduct the interpolation and extrapolation experiments. Here, all data for each subject was sorted with respect to DBP/SBP value, then 10 mmHg bins for MIMIC-III and 5 mmHg bins for Bio-Z were segmented where each was iteratively held out as a testing set while the remaining bins were used for training. Bin sizes were set to 10 mmHg for MIMIC-III due to the range of BP values for each subject being approximately twice as large on average compared to those in the Bio-Z dataset. These experiments evaluate the quality by which each model can estimate BP values that were not

TABLE V
INTERPOLATION AND EXTRAPOLATION RESULTS FOR BIO-Z AND MIMIC-III DATASETS

Dataset	Model	DBP			SBP		
		RMSE (mmHg)	$\overline{\Delta RMSE}$ (mmHg)	r	RMSE (mmHg)	$\overline{\Delta RMSE}$ (mmHg)	r
MIMIC-III	EF-MLP	8.32 ± 10.38	-	0.64	12.80 ± 16.26	-	0.77
	Bi-GRU+Attn	10.15 ± 11.92	-	0.58	13.96 ± 6.76	-	0.72
	DDGA+Bi-GRU+Attn	9.33 ± 11.25	-0.75	0.65	13.45 ± 16.69	-0.44	0.74
	CNN+Bi-GRU+Attn	11.19 ± 12.76	-	0.56	17.28 ± 20.72	-	0.70
	DDGA+CNN+Bi-GRU+Attn	10.39 ± 12.19	-0.32	0.61	17.19 ± 19.84	-0.13	0.72
	Transformer	10.45 ± 11.94	-	0.22	14.72 ± 18.04	-	0.66
Bio-Z	DDGA+Transformer	10.26 ± 11.71	-0.15	0.26	14.59 ± 17.96	-0.09	0.67
	EF-MLP	8.34 ± 12.50	-	0.60	20.22 ± 35.87	-	0.41
	Bi-GRU+Attn	6.70 ± 12.20	-	0.63	9.17 ± 18.51	-	0.70
	DDGA+Bi-GRU+Attn	6.16 ± 12.28	-0.53	0.69	8.81 ± 18.50	-0.34	0.73
	CNN+Bi-GRU+Attn	6.18 ± 12.49	-	0.71	8.80 ± 18.76	-	0.74
	DDGA+CNN+Bi-GRU+Attn	5.90 ± 12.29	-0.34	0.72	8.23 ± 18.62	-0.62	0.77
Bio-Z	Transformer	5.80 ± 11.84	-	0.69	8.45 ± 17.70	-	0.71
	DDGA+Transformer	5.56 ± 11.71	-0.23	0.72	8.25 ± 17.71	-0.20	0.72

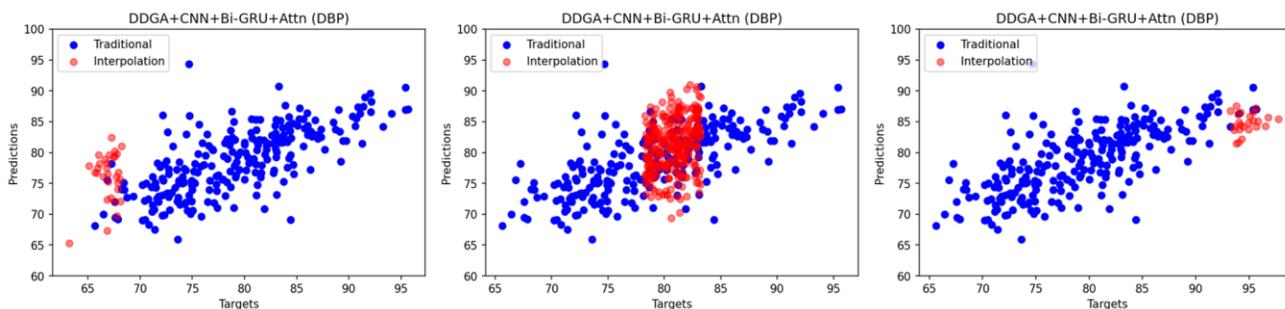


Fig. 5. DBP estimation performance for a sample patient, comparing the test predictions by DDGA from the interpolation and extrapolation test in red and the predictions from the traditional training/testing approach by DDGA in the previous experiments from Section IV-D in blue.

previously seen in training. For real-world scenarios, it is unlikely that the complete range of plausible BP values can be represented in training, especially for those that are solely associated with extreme health conditions. However, it is critical that these abnormal BP instances are captured to ensure the early detection of CVDs and life-threatening cardiac events.

In Table V, we show DBP and SBP performance for the estimation model that leverages extracted features, each baseline architecture, and for the improvement observed with DDGA for both datasets. Prediction performance was evaluated holistically for all test bins for all subjects/patients and also with respect to the average improvement per each subject/patient after applying DDGA to each baseline model, similar to our previous experiments. The first baseline using extracted features achieved the best performance for the MIMIC-III dataset but underperformed for the Bio-Z dataset. This further highlights the impact of DDGA to improve model generalizability when the training dataset size and variance is not abundant. We observe overall improvements to all DL baseline architectures that analyze raw waveforms by up to 8.1% for each dataset. Therefore, proving DDGA’s ability to generalize DL models to ranges of BP that were not previously seen in training. In Fig. 5 we show an example of the predicted bin of DBP values held out for various iterations of this interpolation and extrapolation experiment for the Bio-Z datasets overlaid over the predictions achieved by the traditional style of training conducted in Section 4.4 where all ranges observed in the test data were seen before in training. We observe strong agreement by the model under both sets of conditions.

V. DISCUSSION

Our experiments demonstrate DDGA improvements to personalized BP estimation for all three baseline architectures, and improve estimation precision for upper and lower quartile ranges of BP that were underrepresented in training. DDGA also provides improvements for interpolation and extrapolation experiments when select ranges of DBP/SBP values were removed from the training dataset and used solely for testing. Thus, DDGA proves its ability to prevent overfitting to dataset bias and enhances personalized model generalizability. Improved BP estimation despite the challenges associated with an imbalanced dataset is due to DDGA’s ability to ensure that the learned feature importance by the DL is meaningful to the underlying physiology and physics of the waveform – with minimal expert intervention. Moreover, for most scenarios, DDGA outperformed the EF-MLP estimation model which analyzes only extracted feature sets pre-defined by domain experts. As aforementioned, hand-crafted feature selection incurs burden to both time and effort while also limiting the scope of analysis by the estimation model. Therefore, the results reflect that granting the estimation model flexibility to analyze the whole physiological waveform while informing it of key physiological events to be more effective for BP estimation tasks. In Fig. 6, we observe generated attention maps which reflect learned feature importance for the given input waveform samples by the model type before and after applying DDGA – each timestep in the waveform is scored where a score of 1 signifies greatest importance and 0 signifies least importance for the instance. Attention maps are generated by taking the

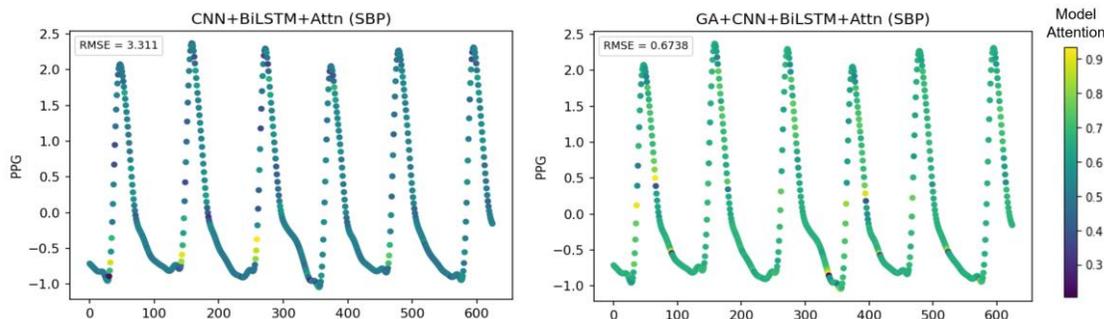


Fig. 6. Comparison of attention maps for baseline models before and after DDGA. We observe high focus to few regions of the input waveform for the baseline models as opposed to after applying DDGA where we observe more distributed attention across the whole waveform and the key fiducial points.

gradient of each output prediction with respect to each timestep of the input as proposed in previous work [47]. For all cases where we observe a remarkable decrease in RMSE, we also observe learned feature importance scores that are more equally distributed to the regions of the waveforms that surround the fiducial points as opposed to the scenario before applying DDGA where model attention scores close to 1 surround only few regions of the input thus limiting the amount of information extracted from it. In future work, the efficacy of these generated attention maps may be evaluated for model interpretability. Comparing these to the identified fiducial points for an input physiological waveform may be used to provide a notion of estimation confidence or quality to the supervising healthcare provider.

Limitations of this work include the lack of analysis on improved model generalizability with DDGA when testing over patients whose collected waveforms were not previously present in training. However, this scenario is distinct from that associated with developing personalized BP estimation models. Particularly, such generalizable models in the health-science domain depend on advanced calibration [48] or domain adaptation [22], [49] solutions which are beyond the scope of this work. Achieving a calibration-free estimation model that is robust to this scenario would be very desirable as it is not practical in the remote health monitoring scenario to collect samples from a newly encountered patient for retraining. In future work, we plan to explore DDGA's generalizability to this setting. While the learned feature sets of DDGA might include some that are common or which were previously proposed for personalized modeling, we hypothesize that remaining features might better capture subject-independent characteristics due to the flexibility granted in the learning process. Furthermore, in future versions of DDGA, we plan to further enhance model generalizability by incorporating information from the learned attention maps into the training process – similar to how it has been done in previous work in the computer vision domain [26]. Particularly, they may be used to provide a notion of uncertainty as we may observe to which morphological characteristics the model chose to prioritize. This information may serve as additional feedback to the optimizer during training.

VI. CONCLUSION

In this work, we proposed a DDGA framework that incorporates minimal domain-knowledge into the automated feature extraction learned by DL models. Effectively, our solution consists of a novel integration that enables researchers to communicate information to the DL architectures through a Guided Attention approach best fit for time-series data types. With only a single template waveform cardiac cycle labelled with key fiducial points, our framework automatically annotates the remaining instances in a training dataset with DTW – which are ultimately used as labels for DL model optimization. Training a BP estimation model to first identify fiducial point locations in an input waveform encourages the model to learn to extract features that are supported by its underlying physiology and physics. Our experiments show DDGA's improvements to personalized BP estimation models composed of three prominent DL architectures. Furthermore, our model increases personalized model generalizability to new ranges of BP not previously seen

in training. DDGA enhances physiological parameter estimation with noninvasive, practical wearable devices that can provide longitudinal measurements throughout patients' daily activities thus providing comprehensive diagnostic information to healthcare providers.

VII. ACKNOWLEDGMENT

This work was supported in part by the National Institute of Health, under grant 1R01EB028106.

REFERENCES

- [1] E. S. Spatz *et al.*, "Phenotypes of hypertensive ambulatory blood pressure patterns: Design and rationale of the ECHORN Hypertension Study," *Ethn. & Dis.*, vol. 29, no. 4, p. 535, 2019.
- [2] 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, doi: 10.1161/HYPERTENSIONAHA.118.11319.
- [3] M. Kikuya *et al.*, "Diagnostic thresholds for ambulatory blood pressure monitoring based on 10-year cardiovascular risk," *Circulation*, vol. 115, no. 16, pp. 2145–2152, 2007.
- [4] P. Verdecchia, D. Clement, R. Fagard, P. Palatini, and G. Parati, "Blood Pressure Monitoring. Task force III: Target-organ damage, morbidity and mortality.," *Blood Press. Monit.*, vol. 4, no. 6, pp. 303–317, 1999.
- [5] E. O'Brien *et al.*, "European Society of Hypertension recommendations for conventional, ambulatory and home blood pressure measurement," *J. Hypertens.*, vol. 21, no. 5, pp. 821–848, 2003.
- [6] Y.-S. Shin, "Identification of blood pressure reflecting personalized traits using bilateral photoplethysmography," *Technol. Heal. Care*, vol. 28, no. S1, pp. 217–227, 2020.
- [7] 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.
- [8] Y. Li, H. Gu, H. Fok, J. Alastruey, and P. Chowienczyk, "Forward and backward pressure waveform morphology in hypertension," *Hypertension*, vol. 69, no. 2, pp. 375–381, 2017.
- [9] O. Singh and R. K. Sunkaria, "Detection of onset, systolic peak and diastolic notch in arterial blood pressure pulses," *Meas. Control*, vol. 50, no. 7–8, pp. 170–176, 2017.
- [10] M. Elgendi *et al.*, "The use of photoplethysmography for assessing hypertension," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–11, 2019.
- [11] E. Finnegan *et al.*, "Pulse arrival time as a surrogate of blood pressure," *Sci. Rep.*, vol. 11, no. 1, pp. 1–21, 2021.
- [12] E. Peralta, J. Lazaro, R. Bailon, V. Marozas, and E. Gil, "Optimal fiducial points for pulse rate variability analysis from forehead and finger photoplethysmographic signals," *Physiol. Meas.*, vol. 40, no. 2, p. 25007, 2019.
- [13] E. E. Gallego Martínez, A. González Mitjans, E. Garea-Llano, M. L. Bringas-Vega, and P. A. Valdes-Sosa, "Automatic Detection of Fiducial Landmarks Toward the Development of an Application for Digitizing the Locations of EEG Electrodes: Occipital Structure Sensor-Based Work," *Front. Neurosci.*, vol. 15, p. 268, 2021.
- [14] 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.
- [15] X. Ding and Y. T. Zhang, "Pulse transit time technique for cuffless nonobtrusive blood pressure measurement: from theory to algorithm," *Biomed. Eng. Lett.*, vol. 9, no. 1, pp. 37–52, Feb. 2019, doi: 10.1007/S13534-019-00096-X.
- [16] W. Chen, T. Kobayashi, S. Ichikawa, Y. Takeuchi, and T. Togawa, "Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration," *Med. Biol. Eng. Comput.*, vol. 38, no. 5, pp. 569–574, 2000, doi: 10.1007/BF02345755.
- [17] 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, doi: 10.1109/TBCAS.2019.2946661.

- [18] G. Slapni Ć 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, doi: 10.3390/S19153420.
- [19] 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, doi: 10.1007/S11082-020-02667-0/TABLES/7.
- [20] C. Sideris, H. Kalantarian, E. Nemat, and M. Sarrafzadeh, "Building continuous arterial blood pressure prediction models using recurrent networks," in *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2016, pp. 1–5.
- [21] H. Song, D. Rajan, J. J. Thiagarajan, and A. Spanias, "Attend and diagnose: Clinical time series analysis using attention models," 2018.
- [22] L. Zhang *et al.*, "Developing personalized models of blood pressure estimation from wearable sensors data using minimally-trained domain adversarial neural networks," in *Machine Learning for Healthcare Conference*, 2020, pp. 97–120.
- [23] 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, doi: 10.3390/S20082338.
- [24] T. Tommasi, N. Patricia, B. Caputo, and T. Tuytelaars, "A deeper look at dataset bias," in *Domain adaptation in computer vision applications*, Springer, 2017, pp. 37–55.
- [25] S. Hong, C. Xiao, T. Ma, H. Li, and J. Sun, "MINA: Multilevel Knowledge-Guided Attention for Modeling Electrocardiography Signals," 2019.
- [26] K. Li, Z. Wu, K.-C. Peng, J. Ernst, and Y. Fu, "Tell Me Where to Look: Guided Attention Inference Network."
- [27] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (pgnn): An application in lake temperature modeling," *arXiv Prepr. arXiv1710.11431*, 2017.
- [28] Y. H. Li, L. N. Harfiya, K. Purwandari, and Y. Der Lin, "Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model," *Sensors 2020, Vol. 20, Page 5606*, vol. 20, no. 19, p. 5606, Sep. 2020, doi: 10.3390/S20195606.
- [29] 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.
- [30] D. Lee *et al.*, "Beat-to-beat continuous blood pressure estimation using bidirectional long short-term memory network," *Sensors*, vol. 21, no. 1, p. 96, 2021.
- [31] J. Esmaelpoor, M. H. Moradi, and A. Kadkhodamohammadi, "A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals," *Comput. Biol. Med.*, vol. 120, p. 103719, May 2020, doi: 10.1016/J.COMPBIOMED.2020.103719.
- [32] X. Xing, Z. Ma, M. Zhang, Y. Zhou, W. Dong, and M. Song, "An Unobtrusive and Calibration-free Blood pressure estimation Method using photoplethysmography and Biometrics," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, 2019.
- [33] T. Fernando, S. Sridharan, S. Denman, and C. Fookes, "Split 'n' Merge Net: A Dynamic Masking Network for Multi-Task Attention," *Pattern Recognit.*, p. 108551, 2022.
- [34] A. E. W. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Sci. data*, vol. 3, no. 1, pp. 1–9, 2016.
- [35] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *Int. J. Biosens. Bioelectron.*, vol. 4, no. 4, p. 195, 2018.
- [36] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Physiol. Meas.*, vol. 33, no. 9, p. 1491, 2012.
- [37] 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.
- [38] F. Miao *et al.*, "A novel continuous blood pressure estimation approach based on data mining techniques," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 6, pp. 1730–1740, 2017.
- [39] P. Su, X. R. Ding, Y. T. Zhang, J. Liu, F. Miao, and N. Zhao, "Long-term Blood Pressure Prediction with Deep Recurrent Neural Networks," *2018 IEEE EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2018*, vol. 2018-Janua, pp. 323–328, May 2017, doi: 10.1109/BHI.2018.8333434.
- [40] Y. Sakurai, C. Faloutsos, and M. Yamamuro, "Stream monitoring under the time warping distance," in *2007 IEEE 23rd International Conference on Data Engineering*, 2007, pp. 1046–1055.
- [41] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intell. Data Anal.*, vol. 11, no. 5, pp. 561–580, 2007.
- [42] M. H. Chowdhury *et al.*, "Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques," *Sensors*, vol. 20, no. 11, p. 3127, 2020.
- [43] C. A. Pearte *et al.*, "Characteristics and baseline clinical predictors of future fatal versus nonfatal coronary heart disease events in older adults: the Cardiovascular Health Study," *Circulation*, vol. 113, no. 18, pp. 2177–2185, 2006.
- [44] V. K. Moitra, C. Guerra, W. T. Linde-Zwirble, and H. Wunsch, "Relationship between ICU length of stay and long-term mortality for elderly ICU survivors," *Crit. Care Med.*, vol. 44, no. 4, p. 655, 2016.
- [45] M. Liu, L.-M. Po, and H. Fu, "Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative," *Int. J. Comput. Theory Eng.*, vol. 9, no. 3, p. 202, 2017.
- [46] W.-H. Lin, F. Chen, Y. Geng, N. Ji, P. Fang, and G. Li, "Towards accurate estimation of cuffless and continuous blood pressure using multi-order derivative and multivariate photoplethysmogram features," *Biomed. Signal Process. Control*, vol. 63, p. 102198, 2021.
- [47] S. A. Siddiqui, D. Mercier, M. Munir, A. Dengel, and S. Ahmed, "Tsviz: Demystification of deep learning models for time-series analysis," *IEEE Access*, vol. 7, pp. 67027–67040, 2019.
- [48] 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.
- [49] J. J. Leitner, P.-H. Chiang, and S. Dey, "Personalized Blood Pressure Estimation Using Photoplethysmography: A Transfer Learning Approach," *IEEE J. Biomed. Heal. Informatics*, 2021.