A Meta-Learning Approach for Fast Personalization of Modality Translation Models in Wearable Physiological Sensing

Ali Akbari, Jonathan Martinez, *Student Member, IEEE*, and Roozbeh Jafari, *Senior Member, IEEE*

***Abstract*—Modality translation grants diagnostic value to wearable devices by translating signals collected from low-power sensors to their highly-interpretable counterparts that are more familiar to healthcare providers. For instance, bio-impedance (Bio-Z) is a conveniently collected modality for measuring physiological parameters but is not highly interpretable. Thus, translating it to a well-known modality such as electrocardiogram (ECG) improves the usability of Bio-Z in wearables. Deep learning solutions are well-suited for this task given complex relationships between modalities generated by distinct processes. However, current algorithms usually train a single model for all users that results in ignoring cross-user variations. Retraining for new users usually requires collecting abundant labeled data, which is challenging in healthcare applications. In this paper, we build a modality translation framework to translate Bio-Z to ECG by learning personalized user information without training several independent architectures. Furthermore, our framework is able to adapt to new users in testing using very few samples. We design a meta-learning framework that contains shared and user-specific parameters to account for user differences while learning from the similarity amongst user signals. In this model, a meta-learner approximated by a neural network learns how to learn user-specific parameters and can efficiently update them in testing. Our experiments show that the proposed model reduces the percentage root mean square difference (PRD) by 41% compared to training a single model for all users and by 36% compared to training independent models for each user. When adapting the model to new users, our model outperforms fine-tuning a pre-trained model through back-propagation by 40% using as few as two new samples in testing.**

***Index Terms*—Modality translation, meta learning, bioimpedance, wearable physiological sensing**

I. Introduction

Modality translation models help advance diagnostic value and interpretability of wearable devices by analyzing the signals collected from their low-power sensors and translating them to their counterpart, highly-interpretable physiological signals that are more familiar to healthcare providers [1], [2]. Typically, deep learning solutions are best-suited for this task due to their superior ability to model non-linear relationships that exist between two signals generated from distinct processes. However, despite the fact that a wearable device will always aim to measure the same physiological behavior, sensor capabilities will vary between users due to differences in demographics, body composition, health state, and physical attributes. Consequently, performance will vary for a generalized translation model and will dramatically suffer when it attempts to analyze wearable signals for a new user whose samples were not included in the training set. Hence, there is a need for training personalized translation models that can leverage the similarities amongst signals collected by the same wearable device for different users, while also being able to distinguish user-specific signal characteristics. Furthermore, these models must be able to adapt to analyzing the signals of new users with a minimum amount of labeled training samples and possibly without any labeled training samples at all. This is critical in healthcare applications because collecting labeled data is often expensive, time consuming, labor-intensive, and burdensome for the users. The aim of this study is to build a framework for modality translation in wearable physiological sensors with the ability to learn personalized models and adapt to new users with very few samples.

We explore fast personalization of modality translation models as they grant diagnostic value to bio-impedance (Bio-Z) signals that are conveniently collected by recently developed wearable devices [3]. Such signals are the measure of electrical impedance of biological cells and tissues with respect to a very small electric current flow [3], [4]. When placed on the chest, variations in the Bio-Z signal will correspond to heart and lung movements, muscle contractions, and blood flow. Although Bio-Z has been used to measure heartrate (HR), respiration rate (RR), pulse transient time (PTT), and blood pressure (BP) [3]–[5], practicing physicians do not yet understand the best features of Bio-Z signals that correlate to each of the subprocesses of the cardiac cycle that cannot be directly inferred by the mentioned metrics. However, through modality translation we can obtain the well-studied electrocardiogram (ECG) counterpart to Bio-Z, which allows us to extract the comprehensive information that a more advanced bio-potential sensor would gather but without needing the additional hardware. Therefore, the number of wearable sensors required to be worn by the user can be reduced, thus providing a more comfortable experience.

Sequence-to-sequence learning solutions are often used to handle the modality translation task [6]. In this approach, a transition function maps the input time-series to the desired

target time-series, where the transition function learns the relationships between these two signals that were generated by distinct processes (e.g., Bio-Z and ECG) [7]. This is commonly achieved with neural network architectures that encode features of the input signal to a compressed representation that is then decoded to the target signal [6]. For the generalizable case, the performance per user is typically enhanced through attention mechanisms that help the translation model focus on the most relevant features of the input [1]. Despite this, there will still be a limit on the success of the model performance as user-specific characteristics exist for both the input and output signals. That is, differences in signal amplitude and morphology will exist per user for both the Bio-Z and ECG signals. Consequently, as aforementioned, models trained on certain users will perform poorly for new users. The intuitive solution to this personalization challenge would be training a separate model for each user [8], which will ignore the shared/similar features of the signals, reduce variance in the training set, and also becomes a space complexity burden. Transfer learning and/or retraining is another solution; retraining algorithms usually require a relatively substantial number of labeled training data to retrain the models [9]. However, gathering such training data is very challenging in most healthcare problems, specifically with wearable devices since it needs users to actively participate in data collection and/or annotation, which becomes burdensome and reduces user adherence. Thus, it is significant to update the models for the new users with very few training samples required from them. Finally, the retraining, specifically for deep learning models, is often achieved through the standard backpropagation algorithm in which the derivative of the output is calculated and propagated through the network to update the parameters [10]. This is a computationally expensive process that is challenging to implement on wearable devices with low-power computational units.

We address the aforementioned gaps by designing a sequence-to-sequence framework based on convolutional neural networks (CNN) augmented with a meta-learning component – a module that learns how to learn user-specific parameters for each particular data. This system can take advantage of learning shared features of the signals from multiple users through its shared feature encoder network, and simultaneously learns a user-specific transition function through a user-specific decoder network. While the weights of the shared feature extraction network are learned through the typical backpropagation method and remain fixed after training, the weights of the user-specific decoder are learned by a meta-learner, which is also approximated by a neural network whose process we refer to as weight calibration. Through the training process, the weight calibration network learns to look at a few samples of a user's data, and updates the weights of the decoder according to user-specific signal characteristics. With this strategy, not only can a single neural network learn a personalized model for each training user, it can also update its weights to handle data for a never before seen user by observing only a few (as low as two samples in this study) or even zero labeled training data of that user in the testing phase. Moreover, learning the model parameters for a new user is achieved through a forward pass rather than gradient backpropagation, which allows it to be used on wearable devices that have low-power computational capabilities. In summary, the contributions of this paper are as follows:

- A meta-learning framework is designed for personalized sequence-to-sequence translation of biomedical signals with the capability of learning both shared and user-specific characteristics of the training data.
- The proposed model can learn personalized deep learning models for each user from a dataset containing multiple users using a single neural network architecture through automatically adjusting parameters of the neural network.
- A methodology is proposed for adapting the models for new users with very few samples of the new data.
- Performance of the proposed methodology is evaluated for bio-impedance to ECG translation through experiments on real data collected from wearable bio-impedance patches to demonstrate the importance of personalized training and model adaptation for new users.

II. Related Work

A. Modality Translation

Successful modality translation models largely depend on the ability to learn complex and non-linear relationships between sensor measurements produced by two distinct underlying processes, which may generally be described as a regression task. For collected wearable signals, this has been achieved on a point regression level by works that correlate photoplethysmogram (PPG) [11], [12] or Bio-Z [8] signal characteristics – such as amplitude, frequency components, PTT, etc. – to estimate heart rate, respiration rate, and blood pressure measurements. However, the task of predicting a whole segment of a separate modality is much more difficult as this requires a translation model to learn several pairwise correlations and generate a totally distinct morphology compared to the source signal. Due to these challenges, deep learning models are often best suited for this task which is typically modeled as a sequence-to-sequence translation problem. Particularly, recurrent and convolutional neural network encoder-decoders have proven to be effective solutions when accomplishing machine translation [13], image captioning [14], and voice dictation [15] applications. When structured into an encoder-decoder architecture, translation is accomplished by dividing translation tasks into feature extraction which is performed by the encoder portion, and into target signal generation, which is accomplished by the decoder. Such approaches have been further extended by attention mechanisms [14], [16].

In addition, there also exist more domain-specific supplements that assist with the modality translation task. For example, text and position embedding can be used to directly relate two words from different languages for machine translation applications [17]. Also, images may be segmented based on detected objects to help determine key words for image captioning [18]. However, such supplements are not as easy to develop for signals collected by wearable devices since they are typically less expressive. Currently, there are very limited previous work that explore machine learning solutions

to handle the modality translation task for signals collected by wearable devices when generating whole segments of a target signal. Discrete cosine transform coefficients of PPG signals collected on the fingertips have been used to infer whole ECG waveforms that belong to the same cardiac cycles [19]. Also, in the direction of attention mechanisms for deep learning, another work has proposed a strategic attention learning framework that enables isolated learning for morphology and amplitude features of an input Bio-Z modality before it is successfully translated to a whole ECG waveform [1]. Although effective solutions, as aforementioned, the notion of personalization for modality translation models is still a great challenge that requires more exploration.

B. Personalization

As aforementioned, personalization of data-dependent learning models is especially critical for wearable devices since collected signals will vary greatly depending on several external factors – sensor placement, consistent motion artifacts, and body type. The intuitive solution would be to train and store a new model for each individual. This has been leveraged in pulse transit time estimation and cuffless blood pressure monitoring with wrist-worn Bio-Z sensors [8], [20]. However, over time, this will require an unmaintainable amount of memory to store each model, a large amount of time to retrain each model before it is able to be used for the new user's data, and depends on the assumption that there will be enough available instances of the new users' data to retrain with or that an adequate amount of time may be spent to collect them. To address the problem of personalizing machine learning models for new user in test time, certain studies leveraged domain adaptation, transfer learning and/or active learning techniques in which a single model is trained for the training data and then it is adapted to the users in the testing phase [21]-[22]. A combination of unsupervised fine-tuning and supervised active learning, which tries to find the most important samples and use them for retraining, was used to personalize activity recognition models with wearable motion sensors [9]. The active learning-based approaches, however, need to have access to an external source to acquire annotation for selected instances, which is not feasible in many health-care applications such as modality translation. That is because in this application annotating the data means collecting the target modality signal in contrast to applications such as human activity recognition where the user can simply provide a single label [23]. Domain adaptation and active learning methods usually require a larger number of data points from the new domain (i.e., the new user) to achieve successful personalization [24]. However, meta-learning-based technique, proposed in this study, can achieve that personalization with only a few samples because during the initial training it learns how to learn the model parameters for a new domain based on domain-specific characteristics as opposed to other techniques that are optimized for a particular domain during the initial training and therefore, they need more data for adapting to new domains.

Mixture of experts and meta-learning are techniques that have been used for building personalized models. A mixture of Gaussian process (GP) experts was designed for personalized detection of cognitive changes in Alzheimer disease [25]. This technique, however, was used for a point regression and it is not applicable to modality transition due to scalability issues. Meta-learning has been used for few-shot learning of new type of outputs in classification tasks [26], [27]. In the current study, we propose a meta-learning-based model that enables few-shot learning as well as personalization for the task of modality translation, which is similar to sequence-to-sequence translation problems and it is basically different from and more challenging than classification tasks in which the outputs are nominal. Our proposed framework learns to update the parameters of the neural network by looking at the overall characteristics of a dataset (consisting of a few samples) instead of single data points. The idea of learning summary of dataset as an important context to guide the machine learning was used to design a conditional variational auto-encoder for image generation [28]. This work used the summary of dataset as an extra input to the decoder. However, our framework updates the parameters of the decoder, which provides more flexibility required for the challenging task of modality translation in which different points of the input signal map distinctly to various points of the target signal.

Our proposed framework can also be compared to a multitask learning framework in which multiple neural networks are optimized at the same time to generate multiple outputs – one per domain/user. This model, however, is not scalable as it requires a comprehensive set of parameters, and is computationally expensive as it needs one neural network per user. Instead, we design an automated process that looks at samples of data of each user and intelligently updates the decoder through a meta-learning process that learns to learn the parameters for each dataset/user.

III. User-specific Variations in Bio-Z and ECG Signals Collected with Wearable Sensors

User-specific variations exist between both the collected Bio-Z and ECG signals. Different individuals' ECG signal posse unique characteristics that impact both the scale and morphology of their typical cardiac cycle characteristics, which are due to both the unique cardiac behaviors that exist within a person and also by the quality of sensing during data collection.

We also observe that although general morphology of the Bio-Z signals is similar, each signal per subject will vary with respect to amplitude, phase, or delay. To gain a better understanding of these variations, we compare the inter and intra-subject variability of ECG signals using cross and auto-correlation. We leverage the maximum of auto-correlation as a measure of intra-subject variability and the maximum of cross-correlation as a measure of inter-subject variability. For two example subjects in our dataset, for ECG signal the auto-correlation (intra-subject variability) is 0.86 and 0.97 respectively while the cross-correlation (inter-subject variability) is 0.81. This shows higher inter-subject variability. Furthermore, for Bio-Z the cross-correlation is 0.65 which shows that the inter-subject variation in Bio-Z signal is more significant than ECG, which has more consistent inter-personal pattern. These types of challenging cases will be best handled with personalized deep learning models.

IV. Methods

In this work, we model the modality translation as a

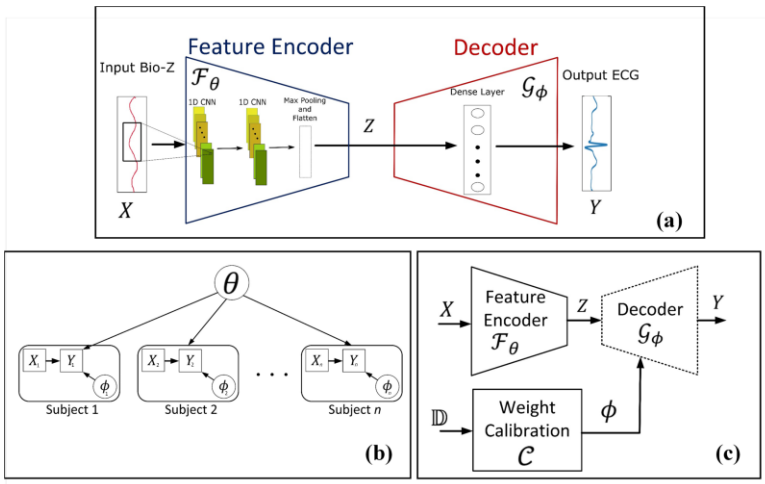

Figure 1. (a) The core modality translator based on CNN encoder and dense decoder; (b) the directed graph model of shared and user specific parameters for modality translation; (c) the full overview of the proposed model

sequence-to-sequence learning problem [6]. As shown in Figure 1-a, we leverage a CNN-based encoder for feature extraction followed by a decoder with a dense layer to translate input Bio-Z to the corresponding ECG signals. The hypothesis is that the low-level features of the signals are transferable between the users while the high-level transition function that contains the relationship between the Bio-Z features and corresponding ECG signal differs from one to another user [29]. Considering this hypothesis, as illustrated in Figure 1-b, our proposed framework consists of shared parameters θ that are the weights of the feature encoder layers and user-specific parameters ϕ that are the weights of the decoder layers. Finally, Figure 1-c shows the full overview of our proposed framework in which a meta-learner, called weight calibration network, is trained to generate the user-specific weights ϕ for the decoder. Similar to the natural way of human learning, the meta-learning module learns to learn good parameters (i.e., weights of the decoder module) given the overall characteristic of each user's data. Before introducing the details of our approach, notation definitions are summarized in Table I to enhance the readability of the article and equations.

A. Meta-learning Framework for Modality Translation

We define the input signal as X and the corresponding output signal as Y. We denote the feature encoder network as $\mathcal{F}_\theta: X \longrightarrow Z$ that maps the input signal to the feature space Z using the shared parameters (i.e., weights) θ. The signal decoder network is shown as $G_\phi: Z \longrightarrow Y$ that maps Z to the desired output modality using the user-specific parameters ϕ. The whole operation of the modality translator can be written as Equation 1, where $\hat{Y}$ is the output of the model, which is an estimation of the target signal Y.

$$\hat{Y} = \mathcal{G}_\phi(\mathcal{F}_\theta(X)) \quad (1)$$

Assuming a dense layer with weights ϕ for the decoder ($\mathcal{G}$), we can rewrite Equation 1 as:

$$\hat{Y} = \phi_w \times \mathcal{F}_\theta(X) + \phi_b \quad (2)$$

where parameters ϕ are divided into the weights ϕ_w and biases ϕ_b. Equation 2 can be implemented as a typical encoder-decoder framework as shown in Figure 1-a, which was also used for modality translation in [1].

TABLE I
TERM DEFINITIONS

Term	Definition
X	The input signal (i.e., Bio-Z in this study)
Y	The output/target signal (i.e., ECG in this study)
Z	The feature space
$\mathcal{F}_\theta$	Feature encoder function that maps inputs to the feature space
θ	Parameters (i.e., weights and biases) of the feature encoder function
G_ϕ	Decoder function that maps features into the target signal
ϕ	Parameters (i.e., weights and biases) of the decoder function
$\hat{Y}$	The output of the decoder which is an estimate of the target signal
$\mathcal{C}$	Meta-learner function that learns to adjust the weights of the decoder
W	Parameters (i.e., weights and biases) of the meta-learner function
$\mathbb{D}$	A subset of each dataset (i.e., each user's data) that is used to train the meta-learner

In our proposed framework, the shared parameters θ are learned through the normal backpropagation algorithm [30], in the training phase, and are kept constant after training. These parameters are trained to extract fundamental spatiotemporal features from the input signal, which are shown to be more transferable in deep neural networks [31]. In fact, the feature encoder module is user-independent in our framework, as opposed to the decoder module that is user-dependent. By sharing parameters for all users, as illustrated in Figure 1-b, we can leverage the similarity between the signal morphology of different subjects to learn a comprehensive feature extraction network. However, the user-specific parameters ϕ need to be adjusted according to the unique signal characteristics of each user. To address this problem, we design a meta-learning framework for learning parameters ϕ from data of each user.

The meta-learner in this study is a neural network itself consisting of CNN and dense layers. As shown in Figure 1-c we call this meta-learner as the weight calibration network since it learns to adjust the weights of the decoder for modality translation – denoted as $\mathcal{C}: \mathbb{D} \longrightarrow \phi$, where $\mathbb{D}$ is a subset of the data of the user. This network will analyze a few samples of each user's data and learn how to generate appropriate weights for the decoder thus incorporating the notion of personalization

Algorithm 1 – Episodic Training

Input: Training data, number of training iteration n_{iter}

1. $i = 0$
2. ***while*** $i < n_{iter}$:
3. ***for*** each dataset **T** in the training data**:** *// each dataset **T** contains the data of one user*
4. $\mathbb{D} \leftarrow n_{\mathbb{D}}$ random samples drawn from **T**
5. $\phi_W , \phi_b = \mathcal{C}(\mathbb{D})$
6. ***for*** j in 1: $n_{\mathbb{V}}$:
7. Draw sample (X_j, Y_j) from **T**
8. $\widehat{Y_j} = \phi_w \times \mathcal{F}_\theta(X_j) + \phi_b$
9. Calculate the loss given Y_j and $\widehat{Y_j}$
10. ***end for***
11. ***end for***
12. Update the weights of $\mathcal{C}$ and $\mathcal{F}$ through backpropagation over the whole batch *// a batch contains $n_{\mathbb{V}}$ samples from each dataset (i.e., user)*
13. $i \mathrel{+}= 1$
14. ***end while***

into the model. The parameters of the $\mathcal{C}$ will be learned through standard backpropagation.

The weight calibration network consists of a feature encoder with convolutional layers and a dense layer, as shown in Figure 2, to map the features to the desired weights ϕ. As mentioned above, the input to this network is a subset $\mathbb{D}$ of each dataset so that the network learns how to generate proper weights for each dataset. To summarize the information in the calibration dataset and to handle the issue of varying amounts of samples in different datasets, we use instance average pooling over all the samples in the output of the last convolution layer in the weight calibration network before passing it to the dense layers for weight generation as illustrated in Figure 2. Therefore, the weight calibration network generates the weights based on the summarized characteristics of each dataset. With this approach, the weight calibration can be achieved by using various number of calibration data (even with a single data point which may be the case for new users).

B. Episodic Training

The proposed architecture depicted in Figure 1-c is learned end-to-end through Algorithm 1. It must be noted that in this study, we treat the data of each user as a separate dataset. Hence, when we refer to a dataset in this section it is all the data of each user. An episodic approach is leveraged for training this framework in order to mimic the condition in the testing time where a few numbers of calibration data is available from test users. In episodic training, as shown in Algorithm 1, at each iteration of the training we randomly select $n_{\mathbb{D}}$ samples from each dataset (i.e., the training data of each user), where $n_{\mathbb{D}}$ is a small number (two or five in this study). These $n_{\mathbb{D}}$ samples are called weight calibration data (denoted as $\mathbb{D}$) and they are fed into the weight calibration network $\mathcal{C}$ to estimate decoder's weights. We also randomly choose $n_{\mathbb{V}}$ samples from each training dataset to pass it through the translator network, estimate translation loss, and update the parameters of the network. In fact, $n_{\mathbb{V}}$ is equivalent to the batch size in standard neural network training, where the average loss is evaluated over all of its samples and one update of the parameters is performed. Adjusting the weights of the decoder through the weight calibration network is achieved by a single forward pass of calibration data $\mathbb{D}$ through it. Regardless of the number of samples of the new data is given to the weight calibration network, it will update the user-specific weights ϕ by feeding all the calibration data to the weight calibration network.

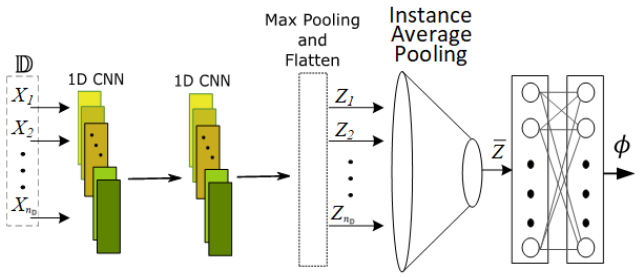

Figure 2. The architecture of weight calibration network

C. Updating the Model in the Testing Phase

To update the model in the testing phase we need to give samples of data of the test user to the weight calibration network to adjust the weights of the decoder and then feed each sample to the encoder-decoder for translation. The number of samples fed to the weight calibration network in the training phase, denoted as $n_{\mathbb{D}}$, could be different from that in the testing phase. In fact, we assume there is much less calibration data available in the testing phase which is a realistic assumption given the challenging nature of data collection in healthcare applications. To take this into account, we propose two variants of our method. In the first case, we assume there is no ECG data (ground truth) available from the test user. In this case, we only feed a few samples of Bio-Z to the weight calibration network. Hence, the weight calibration network adjusts the weights of the decoder in an unsupervised manner merely based on the characteristics of the input signal. In the second variant, we assume very few ECG samples (i.e., target signal) are available from the new user. This number could be as low as one or two cycles in certain applications. In this variant, we concatenate ECG and Bio-Z signals and feed both as the input to the weight calibration network. In this case, the weight calibration network learns a more optimal set of weights for the decoder since it has access to the underlying relationship between the input and output. In this case, the weight calibration network is able to adjust appropriate weights for the decoder with very few samples (one or two) because it has already learned how to learn good weights for each type of Bio-Z–ECG relationship, as opposed to retraining or fine-tuning paradigms that need more labeled samples to resolve the optimization-based training of the weights.

D. Data Preprocessing and Neural Network Architecture

In this section we introduce the details of the signals and architecture of the neural networks used in this study as shown in Figure 1-c. ECG and Bio-Z signals are both collected from the chest similar to the system described in [4]. Each ECG instance is isolated per heart beat centered on the peak value – typically the R-peak but also in a few cases the T-wave. Bio-Z instances are then segmented with the starting point being the time-stamp of the peak value detected in the ECG waveform.

TABLE II
DETAILS OF FEATURE ENCODER NETWORK

Layer	Details
1D Convolution_1	64 Kernels, Kernel size = 3, Activation = ReLU
1D Convolution_2	64 Kernels, Kernel size = 3, Activation = ReLU
Max-Pooling	Pooling size = 2
Dense	Number of neurons = 64

TABLE III
DETAILS OF WEIGHT CALIBRATION NETWORK

Layer	Details
Instance Average Pooling	-
Dense_1	Number of neurons = 64
Dense_2	Number of neurons = 64
Dense_3	Number of neurons = 64×400 + 400

We start the segment from this time-stamp since the Bio-Z behavior response is expected to be somewhat delayed after the heart beats. Each ECG and Bio-Z signal segment is 400 samples in length (approximately 1.2 seconds given sampling rate of 333 Hz in this study) since this was empirically determined to be the best suited length that included all of the cardiac cycle information necessary to successfully accomplish modality translation.

We followed the same CNN-based architecture as [1] for evaluation purposes. This architecture includes one CNN with two convolutional and max-pooling layers with the details explained in Table II. For tuning the hyperparameters, we performed a grid search approach in which we changed the learning rate (from 0.005 to 0.05 in steps of 0.005), also added the number of layers gradually from one layer. The number of kernels and kernel size was fixed based on the signal characteristics (sampling rate and length of the signals). To select the best architecture for the encoder network, we tested different number of CNN layers as well as replacing it with a fully LSTM network; we did not observe a significant difference in the performance, therefore, we selected the two-layer CNN for the encoder network. It is worth mentioning that this framework can easily be adopted by other applications including point regression or sequence-to-sequence translation by changing the architecture of the encoder and/or decoder. The decoder contains a single dense layer as shown in Equation 1, where the number of weights ϕ_W is equal to the *length of features* × *length of output signal*. Given the length of the features in this study, which is 64, and the length of the output ECG is 400, the ϕ_W is a 64×400 matrix and the bias vector ϕ_b is of length 400.

The weight calibration network includes a CNN-based feature extraction component with the same architecture as shown in Table II. The last layer of this CNN is followed by an instance average pooling layer that takes the average of features over all calibration samples. The output of this average pooling is fed to three dense layers with the details shown in Table III. The output of this network has size 64×400+400 corresponding to ϕ_W and ϕ_b respectively.

V. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of the proposed methodology in designing the user-specific model as well as its capability in adapting the trained models to new users. Our focus is on the task of modality translation for wearables, where we investigate translating Bio-Z to ECG signals. In the following sections, we first introduce the details of the experiments and the dataset collected with wearable, wirelessly coupled Bio-Z patches. Although Bio-Z is a convenient modality for measuring RR, HR, and BP, translating it to a well-known and highly-interpretable signal such as ECG is significant for clinical purposes. Similar to many other healthcare applications, personalization and calibration for new users are important challenges that are addressed in this paper. To show the effectiveness of the proposed model regarding training personalized models, we compare its performance to the standard approaches of training a single model for all the users as well as training separate models for each user. Afterwards, we investigate the capability of this methodology in terms of learning the translation task for a new user with very few training samples and compare it to the current wearable sensing modality translation methods. We leverage percentage root mean square difference (PRD), weighted diagnostic distortion (WDD) [32], and average Pearson correlation between the model's prediction and the ground truth ECG signals to quantify how well the machine learning models generate ECG. The PRD is calculated by dividing the typical RMSE by the average L2-norm of the signal using Equation 3 to facilitate interpretation:

$$PRD = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2}} \quad (3)$$

where n is the total number of test samples. A PRD close to 1 means that the prediction is totally off and values close to 0 mean that the predictions are aligned with the ground truth ECG. The PRD may also be considered as the percentage error of amplitude with respect to the range of the signals. PRD metrics primarily capture amplitude precision of a prediction throughout the whole sequence while correlation primarily evaluates the predicted waveform morphology. Unlike other conventional distortion measures, WDD contains direct diagnostic information as it measures the difference between clinically acceptable features derived from PQRST complex. WDD is a number between 0-100 where smaller numbers show more similarity and larger numbers show bigger difference between the reference and the reconstructed ECG signals. The difference is measured by using certain duration and amplitude features extracted from different critical points on PQRST complex. For more information about calculating WDD please see [32].

A. Data Collection

Bio-Z and ECG signals were collected from ten healthy subjects for a duration of twenty minutes, under IRB approval IRB2017-0086D by Texas A&M University, using the same procedures as [4]. With respect to Bio-Z, a transmitter placed underneath the left armpit injects a harmless AC current into the side of a subject while a receiver electrode is placed across the chest to collect, amplify, low-pass filter, process, and compose the Bio-

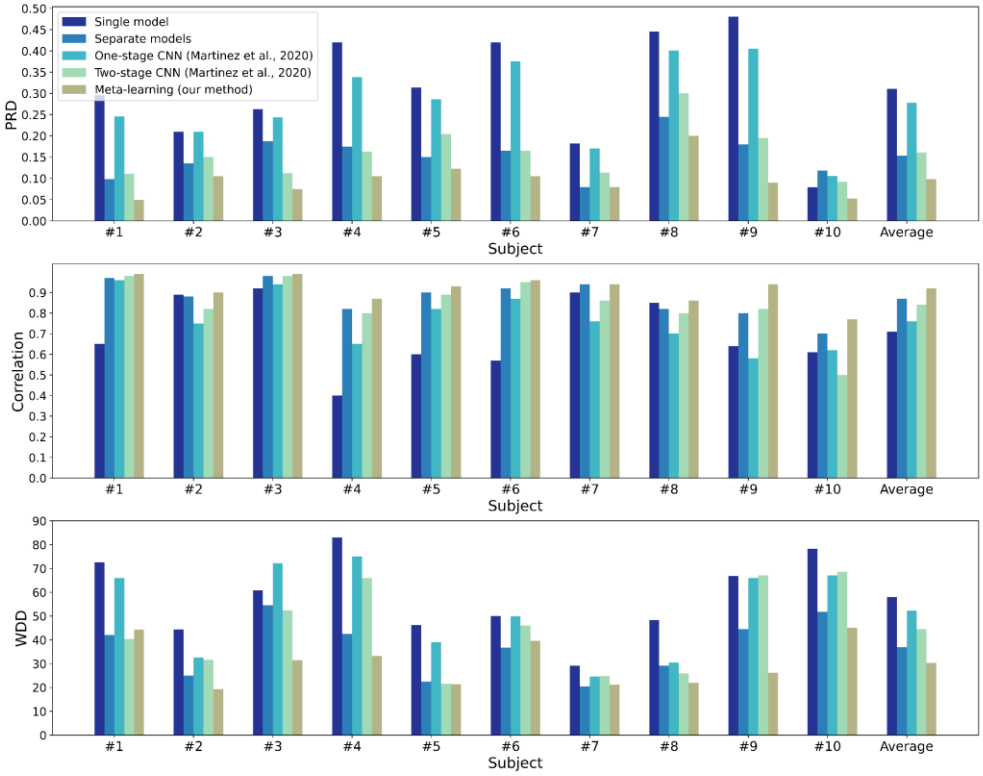

Figure 3.PRD (top), correlation (middle), and WDD (bottom) of Bio-Z to ECG translation when train and test subjects are the same

Z signal. B oth ECG and Bio-Z signals are down sampled to 333 Hz to assist with computation before isolating instances to 400 timestep in length (approximately 1.2 seconds). Ultimately, our dataset encompasses input and output signal pairings that relate to the same cardiac cycle. This yielded a dataset of approximately 13.7k cardiac cycles in total with approximately 1,370 samples per user on average. However, the maximum number of samples per user is 1792 while the minimum number of samples per user is 786. The training process for our dataset took ~3 hours on a server with one NVIDIA GeForce GTX 1080 Ti GPU, two Intel® Xeon® processor E5-2650, and 16x16 GB memory.

B. Personalized Training Performance

In this section, we aim to compare the performance of the personalized model, proposed in this study, with the common approaches of training a single model for all the users as well as training separate models for each user in the training data. In this setting, our proposed method will be trained over all subjects' data with the aforementioned training procedure and the weight calibration will be leveraged in testing as it will update its transition function parameters depending on the subject being analyzed in test time. The need for using personalized models in healthcare applications is critical due to differences in physiological baselines of different users. This comparison reveals the importance of designing such machine learning models that consider these user differences. As mentioned in Section IV, our proposed model learns how to learn user-specific parameters while it also leverages all the training data to learn shared feature extraction parameters. Hence, we expect it to outperform the previous approach of training a single model for all users, which ignores user differences, as well as the approach of training separate models for each user in which the shared features of the signals are completely ignored.

Figure 3 summarizes the PRD, WDD, and correlation between the translated ECG and ground truth. For this experiment, we assume the training and testing subjects are the same. The data is divided into 80% for training, 10% for validation, and 10% for testing. In the test time, we only used two samples of the test data, in an unsupervised manner, for adjusting the weights. That is, we feed only the input Bio-Z from two cardiac cycles before each test sample as the input to the weight calibration method. We perform comparison between our personalized model and:

- training a single model for all the users with the same architecture as shown in Figure 1-c. In this baseline method, we put the data of all users in one dataset instead of considering them in different datasets;
- training a separate model for each user with merely their own data using the same structure as Figure 1-c;

using the current modality translation method based on single stage and two stage translation proposed in [1]. In this approach, one model is trained for all the subjects. For the single stage approach, the encoder-decoder architecture is the same as that described in Figure 1-a, where Bio-Z instances are fed into the network to translate it into the target ECG waveform of the corresponding cardiac cycle. For the two-stage approach, two encoder-decoders of the architecture described in Figure 1-a are cascaded as a Morphology Translation Model and an Amplitude Correction Model. That is, amplitude features of the input Bio-Z instance are filtered out through independent scaling before it is passed into the first stage of the proposed framework. In this way, the output of the first stage is the normalized morphology of the corresponding ECG signal.

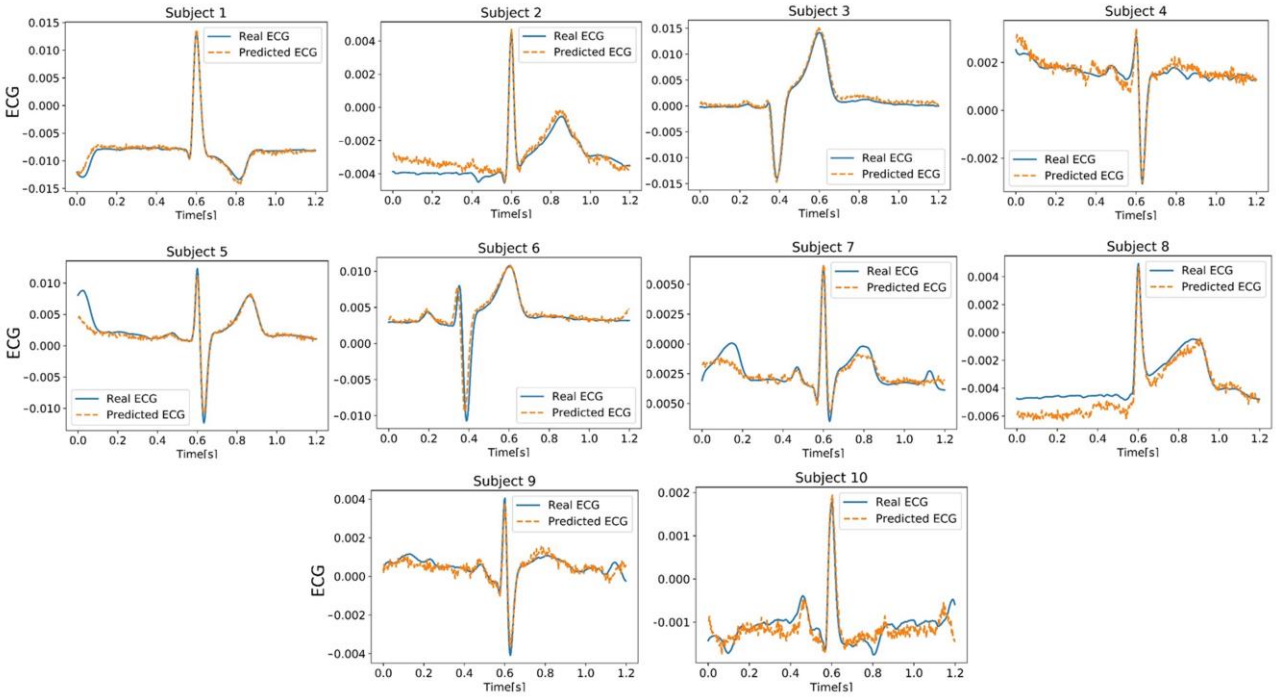

Figure 4. Samples of translated ECG signals by our personalized model when train and test subjects are the same

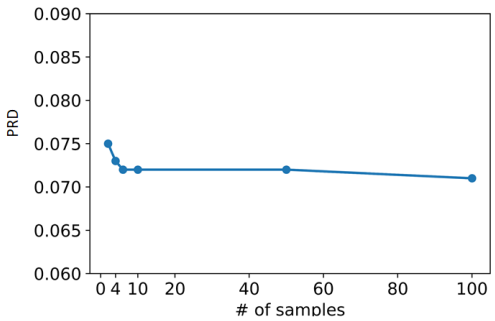

Figure 5. PRD vs. number of samples in case the training and testing subjects are the same.

Then, the raw Bio-Z instance with the actual amplitude (before independent scaling) is concatenated column-wise to the learned morphology of the first stage model's output before they are jointly pass ed into the second stage together to adjust the learned morphology to its correct scale. Ultimately, the output of the second stage is the final translation of the ECG signal.

As Figure 3 shows, our proposed personalized model outperforms the single model as well as the single and two stage networks in [1] by 0.21 (69%), 0.18 (65%), 0.06 (41%) less PRD and 27.5 (47%), 21.9 (42%), and 14.1 (32%) less WDD on average, respectively. Small WDD values show that most of the diagnostic features of the ECG are reconstructed successfully with small distortion compared to the reference ECG. The reason for this improvement is the fact that the proposed model learns user-specific parameters based on the summary of the data of each user as explained in Section IV-A. This helps the system to adjust the parameters of the machine learning model to the data of each user in the test time. This is similar to training an expert model for each user and then intelligently select the right expert in the testing phase by using a few samples of the test user (i.e., two samples in Figure 3}). However, in other three approaches, only one model is used for all the subjects without considering cross-user differences. As shown in Figure 3, our proposed method outperforms the approach of training separate models by 0.05 (36%) less PRD and 6.6 (18%) too. The reason behind this is that our proposed model can also learn shared features of the signals from multiple users, through shared feature encoder parameters θ, whereas training a separate model for each user ignores this shared information. This improvement in the performance proves the importance of training personalized models that can learn both shared and user-specific characteristics. This is an important challenge not only in the application of modality translation, but also in any application that require machine learning for health data.

In addition, per Figure 3, the correlation between the prediction of our model and the ground truth ECG is 0.08 (9%) higher than the two stage network [1], 0.16 (21%) more than the single stage network [1], 0.05 (6%) more than using separate models, and 0.21 (30%) higher than using a single model. This improvement in correlation, however, is less significant than that in PRD, which shows that the most improvement is achieved in estimating the amplitude of the ECG signal rather than the morphology. In this specific problem, since the morphology of the ECG signal is consistent to some extent among our healthy subjects, learning it is easier even with a single model. On the other hand, the amplitude of various parts of the signal (e.g., R-peak) could be very different among even healthy users and that is where the personalized model makes the great impact on the performance. That is why the two-stage network in [1] also performs relatively well as it is designed for correcting the amplitude of the prediction based on the attention mechanism.

We visualize samples of the predicted and ground truth ECG of each user in Figure 4. It can be seen in the figure that, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3105055, IEEE Journal of Biomedical and Health Informatics

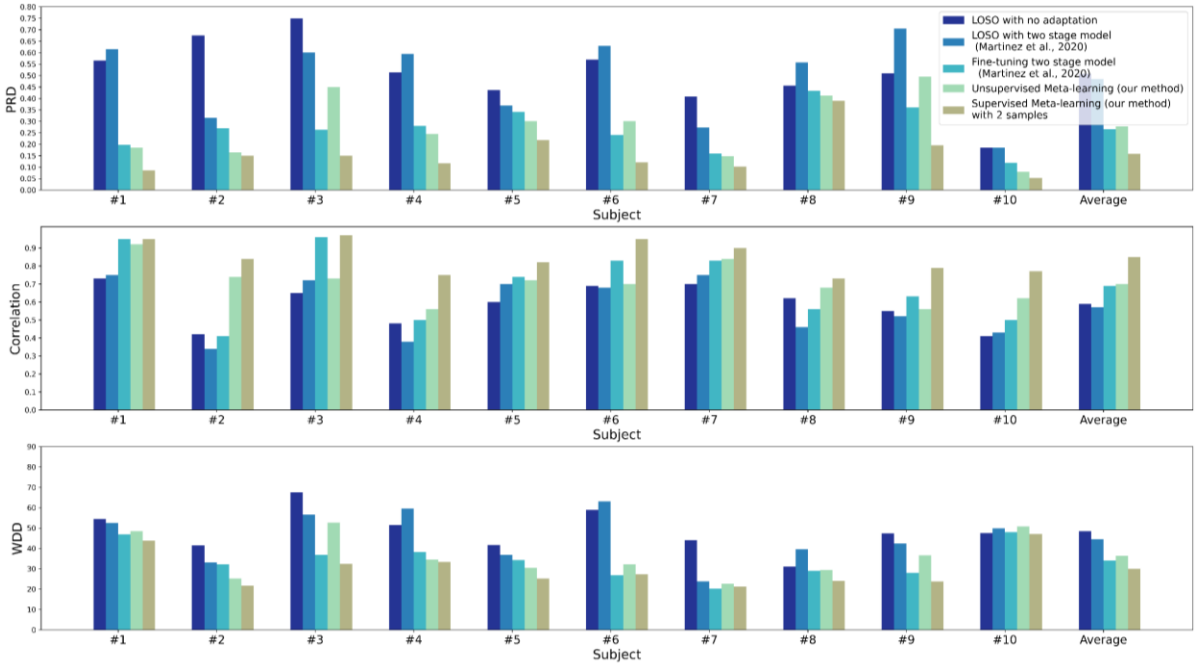

Figure 6. PRD (top), correlation (middle), and WDD (bottom) of Bio-Z to ECG translation for leave one subject out (LOSO) and model adaptation to new subject

predicted ECG follows the ground truth very closely, especially around the R-peak which is one of the most important fiducial points of ECG signal. Most of the errors happen around the parts of the signal where there is no substantial variation. With the proposed methodology, the ECG signals can be acquired for a user wearing a bio-impedance patch without them required to place ECG electrodes. This methodology can be adapted to other applications such as translating bio-impedance to continuous blood pressure, and/or PPG signals too.

We also investigate the amount of data required from the test subjects to adjust weights of the decoder through the weight calibration network. Figure 5 shows the PRD versus the amount of data used from a test subject as the input to the weight calibration neural network averaged over all the subjects. As the figure shows, the proposed model can adapt the weights for the test users with as low as two samples as long as the training and testing subjects are the same. In this case, increasing the number of samples does not make a significant improvement in the performance of the model since the model can easily select the right parameters for the test user. This is different in the case of learning the translation model for a new user (i.e., a user whose data were not available in the training set). In the next section, we will investigate the case of adapting a model trained on certain users for a new user in the testing phase.

C. Model Adaptation for New Users

In most real-world healthcare scenarios, a machine learning model is expected to be used by a new user whose labeled data were not available in the training phase. In this practical case, the pretrained machine learning models usually perform poorly due to differences in the distribution of the new data. This issue raises the need for adjusting/calibrating the models upon characteristics of the new data. In this section, we demonstrate the performance of our proposed model when it is faced data of a new user in the testing time.

We investigate two variants of our proposed framework and compare it to a few other methods, where the PRD and correlations can be found in Figure 6. The first variant of our model assumes that there is no ECG reading available from the new user. Thus, it only uses the Bio-Z data as the input to the weight calibration module. Herein, as the input to the weight calibration module we use two cardiac cycles before the current test sample that is given to the modality translator for estimating its corresponding ECG. This can be considered as a zero-shot learning of the task for new users. This method is referred to as an unsupervised version of our method in Figure 6. In the second variant, we assume that there are a few ECG samples available from the new user. In Figure 6, the assumption is that only two ECG samples are available from the new user. This ECG can be collected by a wearable device such as Apple Watch™ series 4+ or Biowatch [33]. Although these devices collect ECG from the wrist, they require the user to place their finger on the watch, which makes it challenging to collect the data for a long period of time; so, it is a realistic assumption to have a few samples from them. Even if the user visits a laboratory to collect calibration data for ECG, it is fair to assume that reducing the time to collect the calibration data is significant since it reduces users' burden as well as data collection expenses. Finally, there are many other healthcare applications in which collecting labeled calibration data is extremely challenging, e.g., cuff-less blood pressure monitoring or glucose monitoring. In these applications, the need for calibration with very few numbers of labeled samples is of paramount importance. This variant is called supervised version of our method in Figure 6. We compare the two variants of our proposed model with:

- using the same architecture as our framework in Figure 1-c with no adaptation for the new user; in this case we use the weights created from the data of old users for a new user. This approach is called leave one subject out (LOSO) with no

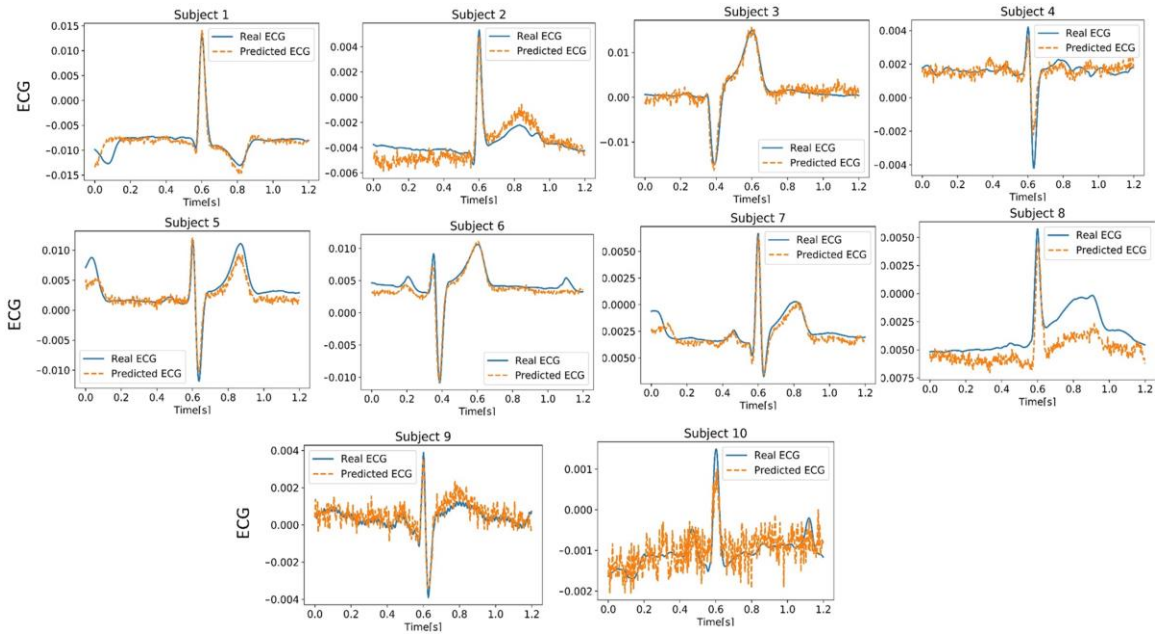

Figure 7. Samples of translated ECG signals in case of adapting our model to new users

adaptation in Figure 6;

- using the best model in [1] with no adaptation where a model is trained on nine subjects and tested on the excluded subject;
- fine-tuning the decoder layer in [1] with two labeled samples via back-propagation (i.e., using both Bio-Z and ECG).

As Figure 6 shows, our proposed framework outperforms the first two methods, which contain no model adaptation, by 0.33 (67%) and 0.32 (66%) less PRD and 18.6 (38.3%) and 14.45 (32.5%) less WDD, respectively. This comparison reveals the importance of model adjustment for new users. Furthermore, our proposed framework outperforms the third case, which is fine-tuning the model in [1] by 0.10 (40%) less PRD and 4.1 (12%) less WDD, which shows the superiority of our framework as it learns how to learn and leverages that knowledge to learn a new task with very few samples. Whereas, in fine-tuning the pre-trained model through typical back-propagation optimization, there is a need for more training data to avoid overfitting. Furthermore, this fine-tuning cannot be done in a unsupervised manner while our proposed framework has that capability. It should also be noted that with this small amount of calibration data (only two samples) training a new model from scratch for a new user is very challenging if not impossible. As Figure 6 shows, using a few ECG samples, even as low as two samples, can improve the PRD by 0.11 (42%) and the WDD by 6.3 (17.4%) compared to unsupervised adaptation. However, an unsupervised version of our proposed model can achieve almost the same performance as supervised fine-tuning of the model [1]. Expectedly by increasing the number of supervised samples, fine-tuning the model in [1] will eventually beat unsupervised version of our method. However, when the amount of labeled data is very small and not enough for fine-tuning the model, our proposed framework is more robust.

Based on Figure 6, the supervised version of our framework achieves 0.16 (23%), 0.28 (49%), and 0.26 (44%) more correlation compared to fine-tuning the model in[1], LOSO[1], and LOSO with no adaptation using our model. We again see that the improvement in PRD is more significant than that in correlation, which shows that estimating the right amplitude is more challenging than the morphology. However, when there is no adaptation (i.e., LOSO models) very low correlation shows that the estimations are not following the morphology closely too. With fine-tuning or model adaptation, the morphology gets corrected faster than the amplitude. Figure 7 visualizes a sample of t he predicted ECG for each user. As the figure shows, similar to Figure 4, the estimation of the ECG around the R-peak is more accurate than the flat parts of the signals. Moreover, we observe that for some subjects (e.g. subjects 8, 9, and 10) the estimated ECG includes tiny spikes, especially around flat regions of the signal. Moreover, we see higher PRD and lower correlation for those subjects in Figure 6. Here an important limitation of the proposed framework, which is the similarity in the feature space, should be discussed. When the new data (i.e., signal of a new subject) can be expressed by the same feature space as the old training data, the proposed model would perform well in adjusting weights of the decoder. However, when there is no similarity in the feature space, the summary of the new dataset (i.e., the output of instance average pooling of features in weight calibration shown in Figure 2) will fall into a new and unknown region of the feature space that is different from the old training data. In this case, the weight calibration module, which is a trained neural network itself, does not know well how to map this summary to the weights of the decoder. As a result, the weight calibration network will not be able to generate proper weights. Thus, when there is a new data with very different characteristics or feature space (e.g., a new user with very different signal characteristics or a data collected by a different sensor), our proposed framework will need more supervised data for learning the new task. This issue can be alleviated by increasing the amount of initial training data. An important property of the proposed framework is that it can learn from distinct datasets collected under different settings. Such a comprehensive training of the weight calibration module on multiple datasets will help the model to better adapt to new data.

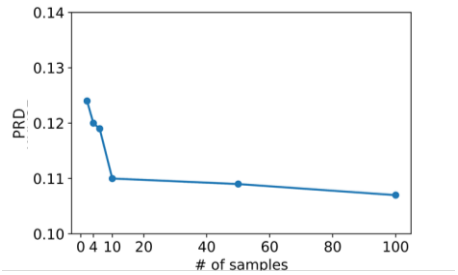

Figure 8. PRD vs. number of samples for adaptation to new user

The comparison in Figure 6 showed that our proposed framework has the capability of learning the translation for a new user with as low as two samples and achieve higher performance compared to other approaches. In Figure 8, we investigate the relationship between performance of this model and the number of labeled samples used for model adaptation. As the figure shows, not surprisingly, increasing the number of labeled calibration samples decreases the PRD consistently. However, with as few as two to ten samples our proposed model can learn the task for the new user with an acceptable accuracy and it gains its most improvement, where there is lesser significant improvement gained after this point. This is achieved by learning the user-specific parameters as the output of another neural network (i.e., weight calibration network). In fact, this framework learns how to learn good parameters for each dataset from its prior knowledge, which facilitates the learning for the new user. Whereas, in typical parameter learning through back-propagation, the best set of parameters are found for a specific data distribution through optimization, but learning the parameters for a new distribution requires solving the optimization problem from scratch.

VI. Discussion

In the problem of modality translation for wearable sensors, training a single generalizable model for all the users leads to ignoring cross-user variations that is present in many healthcare problems due to physical and physiological differences while training separate independent models for each user discards the invaluable information that is common between different users. In the specific problem of translating Bio-Z to ECG, the general morphology of the signals is somewhat similar between different users, but, the amplitude and the patterns of particular fiducial points may vary. In this study, we showed that such a personalized model with both shared and user-specific trainable parameters can achieve superior performance. The key to achieve such a personalized model is to learn how to learn appropriate parameters for each distribution of data through a meta-learner, which is estimated by a separate neural network - weight calibration network. This approach, known as meta-learning is in contrast to the standard training through backpropagation, which finds the best parameters for a specific dataset through solving an optimization problem. In optimization-based approaches, a new optimization problem should be solved for a new setting and this requires a significant amount of training data. However, the meta-learning based framework learns how to learn the best parameters for each data from its prior experience and with that, it can adjust the parameters of the model in the testing time with very few samples through a single forward pass. This is important for wearable solutions in which the models should often work on embedded systems with low computational power. In the testing phase, Bio-Z signal of each heartbeat is processed in 0.08s on average. It should be noted that the time needed for model adaptation is the same as the test phase since it is performed through simple feedforward operations rather than backpropagation-based optimization. Therefore, our proposed model can be used in real-time applications. In addition, given simple nature of deep learning modules used in this model (i.e., convolutional and dense layers) it can be implemented on wearables with limited computational resources and can be used in remote health monitoring applications.

Furthermore, since in the meta-learning based framework, the weight calibration module learns how to generate the best parameters for each dataset, it can easily update the model for the data of a new user. This is vital as it reduces the need for collecting training data for each new user, which is burdensome in most healthcare applications. When the weight calibration module is trained properly, not only can it generate weights for the training data (and data similar to the training data), but it can also estimate weights for previously unseen data under the assumption that the new dataset belongs to the same feature space (or modality) as the previous training data. Since the ECG and Bio-Z are time-dependent, and they can vary over long period of time due to personal changes, there might be a need for our algorithm to be updated periodically for an individual with a few new samples of their signals.

The minimum number of samples required for successful personalization depends on the inter-subject variability and the level of similarity of the new domain (i.e., new user) to training domains. The more various domains used for the initial training, the higher will be the likelihood of successful retraining with few samples for new individuals. An important future direction is to design a model to learn how much data is required from a new domain for the personalization to be successful.

Although the meta-learning approach enables fast learning of personalized models for multiple users via a single neural network, there is an important consideration for using it to adapt the models to new users. The weight calibration module is essentially a neural network that learns to estimate the weights for the decoder in modality translation. Therefore, this module learns to map the summary of a dataset to a large number of weights required for the decoder in modality translation. For instance, in our model, the weight calibration module estimates $400 \times 64 + 400 = 26000$ parameters for the decoder. Similar to any deep learning model, although the deep learning module learns to generalize from the training dataset to unseen testing instances, the accuracy of this module is limited when the distribution of the new data in the testing time is significantly different from the training data. As a result, the proposed meta-learning framework will not be able to adapt the model accurately when the distribution of the data of a new user in the test time is significantly different from the training users and in this case the performance degrades drastically. In this case, a larger number of labeled samples from the new user is required to retrain both the encoder and decoder layers. This is because in addition to the weight calibration module, the encoder will fail to extract meaningful features when the new data includes novel and unseen patterns. A future study can

investigate what level of difference in the signals can be handled by this model and how much retraining would be enough for the encoder to learn the patterns of the new data.

VII. Conclusion

In this paper, we proposed a meta-learning–based framework with the capability of learning personalized modality translation models, and few-shot adaptation for new users. Our proposed framework has the capability of learning how to learn appropriate parameters for a personalized neural network given a few samples of each dataset. Although the proposed framework learns to adjust the user-specific parameters, it has the capability of learning shared features from the data of a different user. Such personalized learning improves translation performance as shown through our experimental results; it also facilitates adaptation of the models to new users with very few calibration samples. Through the experiments where we reconstruct ECG signals from Bio-Z collected by wearables, we showed that our proposed personalized model outperforms the approach of training a single model for all users as well as the approach of training a separate model for each user. Moreover, with as low as two training sample from the new user the proposed model improves the PRD by 40% compared to fine-tuning a pre-trained model through standard back-propagation algorithm. The proposed technique can be applied to other bio-signals such as PPG that have a consistent morphology. The proposed framework is significant since it enables researchers to extract a highly-interpretable ECG signal from a novel Bio-Z signal which is conveniently collected with low-power sensors and it enables new paradigms in designing personalized and adaptable machine learning.

Acknowledgment

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

References

[1] J. Martinez, A. Akbari, K. Sel, and R. Jafari, “Strategic Attention Learning for Modality Translation,” 2020.

[2] M. M. Baig, H. Gholamhosseini, and M. J. Connolly, “A comprehensive survey of wearable and wireless ECG monitoring systems for older adults,” *Med. Biol. Eng. Comput.*, vol. 51, no. 5, pp. 485–495, 2013.

[3] B. Ibrahim, D. Mrugala, and R. Jafari, “Effects of Sensor Placement Relative to the Arterial Sites for Accurate Cuffless Blood Pressure Measurement,” 2019.

[4] K. K. Sel, J. Zhao, B. Ibrahim, and R. Jafari, “Measurement of Chest Physiological Signals using Wirelessly Coupled Bio-Impedance Patches,” 2019.

[5] M.-C. Cho, J.-Y. Kim, and S. Cho, “A bio-impedance measurement system for portable monitoring of heart rate and pulse wave velocity using small body area,” in *2009 IEEE International Symposium on Circuits and Systems*, 2009, pp. 3106–3109.

[6] G. Wang, “Deep Text-to-Speech System with Seq2Seq Model,” *arXiv Prepr. arXiv1903.07398*, 2019.

[7] A. Ghods and D. J. Cook, “Activity2Vec: Learning ADL Embeddings from Sensor Data with a Sequence-to-Sequence Model,” *arXiv Prepr. arXiv1907.05597*, 2019.

[8] 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.*, vol. 13, no. 6, pp. 1723–1735, 2019.

[9] A. Akbari and R. Jafari, “Personalizing Activity Recognition Models with Quantifying Different Types of Uncertainty Using Wearable Sensors,” *IEEE Trans. Biomed. Eng.*, 2020.

[10] M. A. H. KHAN, N. Roy, and A. Misra, “Scaling human activity recognition via deep learning-based domain adaptation,” 2018.

[11] A. Chandrasekhar, M. Yavarimanesh, K. Natarajan, J.-O. Hahn, and R. Mukkamala, “PPG Sensor Contact Pressure Should be Taken into Account for Cuff-Less Blood Pressure Measurement,” *IEEE Trans. Biomed. Eng.*, 2020.

[12] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, “Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks,” *Sensors*, vol. 19, no. 14, p. 3079, 2019.

[13] I. Sutskever, O. Vinyals, and Q. V Le, “Sequence to sequence learning with neural networks,” in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[14] K. Xu *et al.*, “Show, attend and tell: Neural image caption generation with visual attention,” in *International conference on machine learning*, 2015, pp. 2048–2057.

[15] C.-C. Chiu *et al.*, “State-of-the-art speech recognition with sequence-to-sequence models,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 4774–4778.

[16] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, “Convolutional sequence to sequence learning,” in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 2017, pp. 1243–1252.

[17] W. Chan, N. Jaitly, Q. Le, and O. Vinyals, “Listen, attend and spell: A neural network for large vocabulary conversational speech recognition,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 4960–4964.

[18] Q. You, H. Jin, Z. Wang, C. Fang, and J. Luo, “Image Captioning With Semantic Attention,” Jun. 2016.

[19] Q. Zhu, X. Tian, C.-W. Wong, and M. Wu, “ECG reconstruction via PPG: A pilot study,” in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 2019, pp. 1–4.

[20] B. Ibrahim, A. Akbari, and R. Jafari, “A novel method for pulse transit time estimation using wrist bio-impedance sensing based on a regression model,” in *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2017, pp. 1–4.

[21] S. Sabato and T. Hess, “Interactive algorithms: from pool to stream,” in *Conference on Learning Theory*, 2016, pp. 1419–1439.

[22] R. Fallahzadeh and H. Ghasemzadeh, “Personalization without user interruption: Boosting activity recognition in new subjects using unlabeled data,” in *Proceedings of the 8th International Conference on Cyber-Physical Systems*, 2017, pp. 293–302.

[23] Z. E. Ashari and H. Ghasemzadeh, “Mindful active learning,” *arXiv Prepr. arXiv1907.12003*, 2019.

[24] A. Akbari and R. Jafari, “Transferring activity recognition models for new wearable sensors with deep generative domain adaptation,” in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*, 2019, pp. 85–96.

[25] O. (Oggi) Rudovic, Y. Utsumi, R. Guerrero, K. Peterson, D. Rueckert, and R. W. Picard, “Meta-Weighted Gaussian Process Experts for Personalized Forecasting of AD Cognitive Changes,” in *Proceedings of the 4th Machine Learning for Healthcare Conference*, 2019, vol. 106, pp. 181–196, [Online]. Available: http://proceedings.mlr.press/v106/rudovic19a.html.

[26] J. Snell, K. Swersky, and R. Zemel, “Prototypical networks for few-shot learning,” in *Advances in neural information processing systems*, 2017, pp. 4077–4087.

[27] X. Ma, Y. Zhao, L. Zhang, Q. Gao, M. Pan, and J. Wang, “Practical Device-Free Gesture Recognition Using WiFi Signals Based on Meta-Learning,” *IEEE Trans. Ind. Informatics*, 2019.

[28] H. Edwards and A. Storkey, “Towards a neural statistician,” *arXiv Prepr. arXiv1606.02185*, 2016.

[29] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” *arXiv Prepr. arXiv1411.1792*, 2014.

[30] R. Hecht-Nielsen, “Theory of the backpropagation neural network,” in *Neural networks for perception*, Elsevier, 1992, pp. 65–93.

[31] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, “Ecg heartbeat classification: A deep transferable representation,” in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 2018.

[32] Y. Zigel, A. Cohen, and A. Katz, “The weighted diagnostic distortion (WDD) measure for ECG signal compression,” *IEEE Trans. Biomed. Eng.*, vol. 47, no. 11, pp. 1422–1430, 2000.

[33] S. S. Thomas, V. Nathan, C. Zong, K. Soundarapandian, X. Shi, and R. Jafari, “BioWatch: A noninvasive wrist-based blood pressure monitor that incorporates training techniques for posture and subject variability,” *IEEE J. Biomed. Heal. informatics*, vol. 20, no. 5, pp. 1291–1300, 2015.