HHS Public Access

Author manuscript

IEEE J Biomed Health Inform. Author manuscript; available in PMC 2025 March 28.

Published in final edited form as:

IEEE J Biomed Health Inform. ; PP: . doi:10.1109/JBHI.2023.3320585.

Variational Autoencoders for Biomedical Signal Morphology Clustering and Noise Detection

Zhale Nowroozilarki [Student Member, IEEE],
Department of Computer Science Engineering, Texas A&M University, College Station, TX, 77840 USA

Bobak J. Mortazavi [Senior Member, IEEE],
Department of Computer Science Engineering, Texas A&M University, College Station, TX, 77840 USA

Roozbeh Jafari [Senior Member, IEEE]
Department of Biomedical Engineering, Electrical and Computer Engineering, Computer Science and Engineering, Texas A&M University, College Station, TX, 77840 USA

Abstract

Accurate estimation of physiological biomarkers using raw waveform data from non-invasive wearable devices requires extensive data preprocessing. An automatic noise detection method in time-series data would offer significant utility for various domains. As data labeling is onerous, having a minimally supervised abnormality detection method for input data, as well as an estimation of the severity of the signal corruptness, is essential. We propose a model-free, time-series biomedical waveform noise detection framework using a Variational Autoencoder coupled with Gaussian Mixture Models, which can detect a range of waveform abnormalities without annotation, providing a confidence metric for each segment. Our technique operates on biomedical signals that exhibit periodicity of heart activities. This framework can be applied to any machine learning or deep learning model as an initial signal validator component. Moreover, the confidence score generated by the proposed framework can be incorporated into different models' optimization to construct confidence-aware modeling. We conduct experiments using dynamic time warping (DTW) distance of segments to validated cardiac cycle morphology. The result confirms that our approach removes noisy cardiac cycles and the remaining signals, classified as clean, exhibit a 59.92% reduction in the standard deviation of DTW distances. Using a dataset of bio-impedance data of 97885 cardiac cycles, we further demonstrate a significant improvement in the downstream task of cuffless blood pressure estimation, with an average reduction of 2.67 mmHg root mean square error (RMSE) of Diastolic Blood pressure and 2.13 mmHg RMSE of systolic blood pressure, with increases of average Pearson correlation of 0.28 and 0.08, with a statistically significant improvement of signal-to-noise ratio respectively in the presence of different synthetic noise sources. This enables burden-free validation of wearable sensor data for downstream biomedical applications.

zhale@tamu.edu .

Keywords

Biomarker Estimation; Biomedical Signal Processing; Deep Learning; Gaussian Mixture Models; Variational Autoencoder

I. Introduction

WITH the ubiquity of wearable devices and Internet of Medical Things (IoMT), digital health monitoring paradigms are proliferating. The pervasiveness of these devices has improved capturing of clinical data, which may lead to personalized prevention of adverse outcomes [1]. Biomedical wearable sensors are electronic devices that can be connected to the human body to collect biomedical signals and monitor activities [2], enabling continuous observation of physiological biomarkers, including in free-living environments. This remote capture can be used for personalized, remote health diagnosis and treatment. For example, non-invasive biomedical signals, such as electrocardiogram (ECG) and photoplethysmogram (PPG), collected from aforementioned devices, provide vital information about the cardiovascular system to be utilized for diagnosis and monitoring of cardiovascular diseases and risk factors [3]. Another biomedical signal modality that can be acquired to analyze cardiovascular function by blood pulsatile sensing with less fallibility to skin color and thickness is bio-impedance (Bio-Z) waveform data [4]. However, most biomedical waveform data modalities are vulnerable to various sources of noise (e.g., ambient noise or motion noise) [5]. As the volume of medical data captured from IoMT devices increases, the quality of data needs to be verified before extracting clinical biomarkers and being used for clinical monitoring. Any reduction in data quality may result in either difficulty in model training and tuning or generate overconfident models, all which lead to predicting incorrect labels [6], [7], [8], [9]. Most advanced methods for data processing and modeling have been implemented using supervised machine learning methods [10]. However, collecting adequate high-quality data for feature extraction and modeling can be expensive, timeconsuming and challenging [11], [12]. More importantly, in the healthcare domain, incorrect prediction can result in serious damage to the patients (e.g., prescribing wrong treatment or not predicting a malignant tumor using radiology images) [13]. Therefore, it is essential to have a framework for quantification of the noise in the input data, without extensive data validation or manual engineering, as part of data preprocessing to enable reliable (IoMT) data modeling. There is a crucial need for an unsupervised model that detects corrupted sequences of biomedical signals with periodic pattern, aiding in noise quantification and preprocessing for data validation as signal volume increases. Data validation is needed to eliminate noise and abnormalities that can lead to a potential decrease in the model's accuracy and reliability [14]. For example, a noisy ECG signal can incorrectly detect/classify heartbeats [15]. Supervised machine learning models can be trained to extract features and assess signal quality [16]. However, labeling noise in the data can be a challenging task. Moreover, classification methods for noise detection do not quantify noise unlike classical approaches like the Signal-to-Noise ratio. While extensive architecture modification and hyperparameter tuning can improve deep learning models for extracting features and learning new relationships in medical data, due to distinct physiological differences between individuals in the various cohorts or inherent variation

in population demographics, deep learning models can be biased to the observed training data without sufficient generalizability [17]. As a result, there is a need to validate the quality of input data before being fed into such models. Moreover, these data validator frameworks should be compatible with different data modalities and should be easily incorporated into different downstream prediction models for detecting the data points, as out-of-distribution samples, either as noisy data or as distributions shift. In this work, we developed an unsupervised method to both distinguish between noisy and clean segments in biomedical signals, with underlying periodicity, and provide a metric for the quality (degree of corruptness) using Variational Autoencoders (VAE), Gaussian Mixture Models (GMM) and Dynamic Time Warping (DTW). We demonstrate the application of the proposed signal quality model using two biomedical signal modalities with a periodic pattern: bioimpedance signals collected internally and ECG signals captured in the MIT-BIH dataset available from PhysioNet [18]. Finally, to show the application of this framework in an end-to-end bio-marker prediction, we demonstrated the impact of this method on a model to estimate blood pressure from cuffless (non-invasive) signals, through increased Pearson correlation and reduced root mean square error (RMSE). The contributions of this work are:

- Providing an unsupervised method for noise detection without the need for any extensive data labeling and validation by domain experts.
- Introduction of a model-free technique that can be applied to any time-series signal modality (e.g., ECG, PPG, Bioimpedance) irrespective of the underlying morphological models for early noise detection.
- Presenting an explainable visualization tool for assessing the quality of biomedical time-series data.
- Providing a probabilistic framework to quantify signal quality.

The remainder of this paper is organized as follows. Section 2 reviews key related works, Section 3 introduces the framework and our methods, while Section 4 discusses the experiments and results. Finally, Section 5 discusses limitations and future work while Section 6 concludes this work.

II. Related Work

Time-series biomedical Noise Detection:

Noise and artifacts in time-series biomedical signals occur from many different sources, including contact noise, baseline drift, or hardware electronic noise, which reduces the ability to extract meaningful information from the data [19]. As a result, noise detection and removal are essential parts of biomedical signal processing which have been explored intensively to find an optimum solution for different data modalities, utilizing techniques such as extended Kalman filter (EKF) [20], [21]. However, these approaches (e.g., methods based on ECG dynamic models (EDM) [22]) do not work if there is a variation in the morphology of ECG segments, such as in patients with arrhythmia, limiting the usability of KF-based approaches [23]. Band-pass filters and Wiener filters are other techniques used for the elimination of different sources of noise with the drawback of not being able to eliminate non-stationary noise [24]. As a result, one would need a comprehensive approach

that can deal with various sources of noise and data modalities and hence a data-driven and model-agnostic method which can work flexibly with different physiological waveform characteristics is preferred.

Deep learning based time-series ”similarity”:

The goal of anomaly detection is to analyze the data to extract meaningful information and further to identify data samples that are representative of valid data points. These methods can then detect outliers deviating from normal observations, likely to represent noise and error in the data [25], [26], [27], [28], [29]. Chauhan et al. utilized a similar approach for ECG anomaly detection where an LSTM (Long Short Term Memory) based model outputted the probability distribution of the prediction error, which was used to detect abnormal behavior [30]. Other types of deep neural networks such as, convolutional neural network-based frameworks, have also been explored for detecting abnormalities in the time-series data [31]. More recently, Deng et al. also used Graph Neural Network to compute the deviation of patterns between different sensors to detect anomalies [32]. Finally, autoencoder-based approaches with different architectures such as a combination of CNN, RNN and fully connected layers, have been proposed in order to identify anomalies [33],[34],[35],[36],[37] where the reconstruction error was used as a proxy for anomaly score. Methods using raw waveform data seek to detect anomalies by clustering morphologies, such as DTW have also been used [38],[39]. As different time points in the time-series can be treated as individual feature, the dimension of this data modality can easily increase throughout the time which makes this approach more susceptible to noise and variation in different data morphology. Thus, other approaches have been proposed to reduce data dimensionality and therefore, better clustering result such as features based on component analysis [40], convolutional and recurrent neural network-based models for feature extraction [41] and integration of temporal reconstruction with k-means clustering [42]. Although the aforementioned frameworks can obtain high accuracy where the data is labeled, there is a need to provide a framework where noisy labels are not present. Additionally, as the timeseries data can be high dimensional, running the approaches based on DTW distances is time-consuming and not efficient. As a result, the goal of this paper is to provide a unique pipeline for biomedical waveform processing which can be combined with any additional model for either regression or classification analysis.

III. Methodology

Our proposed framework, comprised of multiple modules, leverages the periodicity of biomedical signals in a minimally supervised validation pipeline to eliminate noise and increase confidence in the quality of biomedical signals for biomarker estimation. Using this framework, the data segments are passed into a convolutional variational autoencoder model that learns the underlying variation of signal morphologies and further translates signals into a lower dimensional space. Using Gaussian Mixture Models, we then find prototypes of different segment morphologies. These prototypes represent differences in clean segments as well as noisy ones. Intuitively, the noisy prototypes have the highest difference in morphology compared to a validated template. As a result, we need to distinguish between these prototypes to remove the cluster containing corrupted segments. Therefore, we utilize

DTW to find the closest and farthest prototypes to the validated template. Using this complete framework, we validate a subset of data without labeling and domain expert analysis and remove the corrupted segments. The proposed framework is presented in Fig. 1. This technique creates a validated subset of data that can be further passed to any deep learning methods for additional biomarker estimation and prediction. In this paper, we demonstrate the effect of using this end-to-end framework on a downstream task of cuffless blood pressure estimation where noisy data easily impedes the model for robust prediction. This section describes the development of each component of the framework to develop the end-to-end signal quality and translation module.

A. Data Segmentation

Fiducial points are the key points on the biomedical waveform data which correspond to the periodic pattern. In the context of cardiac waveform data, this cardiac pattern captures heart activity. As a result, to segment the waveform data into distinct cardiac cycles, we used the maximum slope points of the Bio-Z waveform data. Similarly, we used QRS complex points of ECG data to segment the data and prepare the input for the data validator framework.

B. Variational Autoencoder Architecture

The primary component of our end-to-end framework is the VAE which provides unsupervised embedding of the signals. An autoencoder is a type of generative model which is comprised of three main components, an encoder, an encoded layer that represents a latent distribution, and the decoder. The encoder's task is to encode the training data into the dimension-reduced latent space. Subsequently, the decoder uses the latent distribution to decode it to the original signal space. The loss function of the autoencoder is defined as follows

$$L_{Reconstruction} = \|x - \hat{x}\| \tag{1}$$

where the model learns to reconstruct ($\hat{x}$) the original input data (x) very closely to minimize the loss. A variational autoencoder combines Bayesian inference with an autoencoder by using a probabilistic measure for the loss function instead of solely relying on the reconstruction error [43]. As a result, VAEs can be considered stochastic generative models, whilst an autoencoder is deterministic. This makes VAE a better solution for anomaly detection, as it is more principled for detecting abnormal distribution in the data [44]. The loss function used in our VAE framework consisted of two components: the reconstruction loss and a regularization term. Using the regularization term, we aimed to satisfy two important properties: 1) continuity, where two close data points in the latent space had similar morphology in the input space accordingly; 2) completeness, where using a chosen distribution, decoded sampled points were close to real values, i.e., meaningful. The VAE implemented in this work is illustrated in Fig. 2. The architectural components of the VAE include 1D convolutional layers as the waveform data used in this paper is 1-Dimensional data. The 1D convolutional layers are coupled with batch normalization layers in order to make the training faster and more stable. Three convolutional layers create the encoder to

extract the higher-level, medium-level and lower-level embedded features for creating the latent space which can be used further for the noise detection algorithm as shown in Fig. 2. The decoder also uses the latent space to reconstruct the original waveforms using the three convolutional layers in the reverse order compared to the encoder. Furthermore, as the reason for the VAE framework is to understand the morphology of each normalized segment, i.e., there is no temporal information about the relation of consecutive cycles, we only use convolutional layers.

C. VAE Formulation

This section presents the formulation of the VAE and how the formulation fits a broader framework. The input data, represented as a set, $X = [x_1, x_2, \ldots, x_n]^T$, contained n number of independent and identically distributed samples (e.g., in this paper, the cardiac cycles from each biomedical waveform can be considered as one individual sample). θ parameters of the model, the latent Z-space was generated from the prior distribution $p_\theta(z)$. The model then sought to learn the joint distribution $p_\theta(x, z) = p_\theta(x \mid z)p_\theta(z)$. In other words, the model learned this joint distribution with a prior distribution $p_\theta(x \mid z)$ where the encoder learns the $p_\theta(z \mid x)$. To learn the joint distribution, we needed to maximize the following likelihood:

$$p_\theta(X) = \int p_\theta(x \mid z)p_\theta(z)dz \tag{2}$$

Variational inference is then used to find an alternative posterior distribution $q_\theta(z \mid x)$, which approximated the $p_\theta(X)$ to make the above integral tractable [45]. To find an approximation as close as possible to the true distribution, the KL-divergence (Kullback–Leibler divergence) of these two distributions then needed to be:

$$D_{KL}(q_\phi(z \mid x) \| p_\theta(z \mid x)) \tag{3}$$

As a result, the variational parameters ϕ were then estimated as:

$$log p_\theta(x) = \mathbb{E}_{q_\phi(Z \mid X)}(log_{p_\theta}(x)) \tag{4}$$

Using Bayes' rule, we replaced $p_\theta(x)$ by $\frac{p_\theta(x, z)}{p_\theta(z \mid x)}$, which resulted in the following:

$$log p_\theta(x) = \mathbb{E}_{q_\phi(z \mid x)}\left(log\left(\frac{p_\theta(x, z)}{p_\theta(z \mid x)}\right)\right) \tag{5}$$

After multiplying and divining the above equation by $q_\phi(z \mid x)$, we have:

$$log_\theta(x) = \mathbb{E}_{q_\phi(z \mid x)}\left(log\left(\frac{p_\theta(x,z)}{q_\phi(z \mid x)}\frac{q_\phi(z \mid x)}{p_\theta(z \mid x)}\right)\right) \tag{6}$$

Finally, the above equation can be written as follows:

$$log_\theta(x) = \mathbb{E}_{q_\phi(z \mid x)}\left(log\frac{p_\theta(x,z)}{q_\phi(z \mid x)} + \mathbb{E}_{q_\phi(z \mid x)}\left[log\frac{q_\phi(z \mid x)}{p_\theta(z \mid x)}\right]\right) \tag{7}$$

where the first term is $L_{\theta,\phi}(x)$ (also called Evidence Lower Bound (ELBO)) and the second term is the KL-divergence. As a result, we have the following:

$$L_{\theta,\phi}(x) = logp_\theta(x) - D_{KL}(q_\phi(z \mid x) \| p_\theta(z \mid x)) \tag{8}$$

By maximizing the $L_{\theta,\phi}(x)$ or more conventionally minimizing the following loss function:

$$L_{vae} = -L_{\theta,\phi}(x) = D_{KL}(q_\phi(z \mid x) \| p_\theta(z \mid x)) - logp_\theta(x) \tag{9}$$

we both maximize the likelihood $p_\theta(x)$ and minimize the KL divergence, making the two distributions as similar as possible. A common choice for the approximation of the posterior distribution is a Gaussian distribution which results in a Gaussian encoder with a common choice of standard normal distribution $N(\theta, I)$ for the prior distribution. In other words:

$$q_\phi(z \mid x) = N\left(z; \mu, \sigma^2\right) \tag{10}$$

Finally, in order to use stochastic gradient descent for model optimization, we need to ensure the objective function is differentiable. As variable z is a random variable, we need to reparametrize it to be deterministic with an additional random variable $\epsilon \sim \mathrm{N}(0, 1)$. Consequently,

$$z = \mu + \sigma \odot \epsilon \tag{11}$$

$$L_{VAE} = L_{Reconstruction} + L_{Regularization} \tag{12}$$

This final loss function (12) was then optimized to train the VAE such that the reconstruction error between the reconstructed waveform and the original waveform is minimized while the distribution of reconstructed segments follows the predefined Gaussian distribution.

D. VAE for Noisy and Clean Waveform Identification

Fig. 3(a) illustrates an example of encoded points presented in a 2D latent space. Each point represents a single input, where in this setting, such input represents a single cardiac cycle. From this encoded space, we sought to define a distance metric, to identify similar signals and dissimilar signals from their Euclidean distance in the encoded space. We define the distance between two points of p and q in an n-dimensional latent space using Euclidean distance.

Applying that distance to our illustrative example above, we can observe two different examples where the points highlighted in green have similar morphology while the points highlighted in orange, because of relatively higher distance in the latent space, have different distinguishable morphology in the input space as well. The morphology for the orange points in Fig. 3(b) and the green points is presented in Fig. 3(c).

E. Gaussian Mixture Models for Latent Space Clustering

A complication with the distance metric, however, is that individual cycles may also be far apart because they represent different physiological stages (e.g., different blood pressures and heart rates captured by wearable IoMT devices). Therefore, the next step of our signal denoising framework is to differentiate distance for noise versus different morphologies/physiologies. We sought to cluster the latent space in order to find the segments which belong to a cluster where the morphologies are mostly uncorrupted, while the points which are further from all clusters would then be identified as be contaminated/noisy in terms of morphology. As the latent space of the autoencoder creates different probability distributions based on the encoded morphologies, Gaussian Mixture Models (GMM) was selected as the clustering method for this work as it distinguishes different Gaussian distributions available in the data. An alternative to the GMM method is K-Means clustering. However, K-Means clusters the data only based on the Euclidean distance of the points in the latent space while GMM provides a probabilistic assignment when selecting the points and assigning them to each cluster which makes this clustering technique more suitable for the problem we are solving in this work. Using GMMs, the latent space was grouped into different clusters, with the idea of having cardiac cycles with different morphologies in different clusters. Furthermore, we defined the closest point to the centroid of each cluster as a representation of that group of cardiac cycles, respectively as the centroid of each cluster is not essentially an actual point in the dataset. As a result, we needed to find the closest point to the centroid (here shown as R_i), as there is a need to map the point back to the raw waveform space: $\forall i \mid i \in k: R_i = argmin D_{j=1}^{n}(x_j, c_i)$ where D is the distance metric defined before, x′s are the data points in the latent space and c's are the centroids of all the clusters.

F. Determining the Number of Clusters

To automatically and accurately define the number of clusters of valid representations and to be able to identify clean and noisy cluster prototypes, we compared different scoring criteria.

First, the Silhouette method was used to interpret and validate the consistency of the data within each cluster and across the subsequent ones [46]. The Silhouette score is a metric between −1 and 1 that measures how similar a point is to the other data points in the current cluster (cohesion) and different it is to the points in other clusters (separation). A higher silhouette score demonstrates better clustering results. The Silhouette score is used to define the range of the candidate number of clusters. Secondly, Bayesian Information Criterion (BIC) score was used to detect the exact number of clusters K, through the maximum likelihood of the n number of data points belonging to K different clusters [47]. Detecting the best number of clusters is crucial as a higher number of clusters usually provides better scores while reducing the size of each cluster, which is not optimal. Additionally, clean segments can also vary in temporal morphology. Therefore, the optimal number of clusters should be controlled to avoid data overfitting as the goal of clean cluster detection is to keep an adequate number of clean segments with natural morphology variation to then be used in the robust biomarker prediction. After identifying the appropriate clustering, we validated segments from the clean clusters with variation in terms of waveform morphology.

G. Cluster Noise Level Detection

While the distance metric provided allows for clustering in the latent space, it does not differentiate clusters of clean signals from clusters with noisy morphologies. In order to automatically find the clusters that contain clean segments, we measured the DTW distance between each cluster's representative cardiac cycle (closest to the centroid) to a single selected clean template, as defined in prior work [39]. This distinguished clusters and the amount of noise each cluster contained.

H. Confidence Metric for Signal Inclusion

The DTW metric provides a method to distinguish validated clusters; however, there is a need for a metric by which we assign a confidence score to each single waveform within the kept cluster. Therefore, we sought to define a confidence metric for keeping signals, the final stage in filtering our signals before providing them to end-to-end prediction frameworks. We sorted all samples within each cluster by their Euclidean distance from the representative of that cluster and normalized these distances between 0 and 1. We defined a weight metric w when the jth segment is the representative of the clean cluster:

$$w_i = 1 - normalized(D(x_i, x_j)) \tag{13}$$

Algorithm 1

VAE-based Framework Algorithm

function VALIDATION($C[], M, GS$)
  $M \leftarrow Size()$
  $N \leftarrow Size(C)$
  for $i \leftarrow 1$ to N **do**

$L_i \leftarrow VAE(C_i)$
end for
$Q \leftarrow GMM(C)$
$P \leftarrow centroid(Q))$
for $k \leftarrow 1$ to M **do**
$d_k \leftarrow DTW(P_k, GS)$
end for
$noisylabel \leftarrow where(d_k = max(d))$
end function

Where the importance of each sample is provided by their similarity to the clean cluster prototype.

I. Data Processing Pipeline and End-to-end Framework

Algorithm 1 defines our end-to-end pipeline implementation for biomarker estimation using the validated gold standard cardiac cycle, where C_i corresponds to each cardiac cycle, L_i corresponds to the latent representation of the cardiac cycle C_i, Q_i corresponds to the GMM label of the cardiac cycle C_i, P contains the prototypes for all the clusters, GS corresponds to the gold standard cardiac cycle template and d contains the DTW distances between the protypes and the GS. The raw waveform data is the input of VAE-based framework for further verification. Using the algorithm explained in the previous section, a subset of cardiac cycles from the validated cluster is used for the blood pressure estimation module. Using this end-to-end framework, we aim to reduce the uncertainty in the data as well as increase the confidence in the robust cuffless BP estimation.

IV. EXPERIMENTS AND RESULTS

We present interpretable visualization based on the proposed framework as well as demonstrate the effect of using the signal validator framework and its generalizability for reducing the variance of DTW distance of signal morphologies to a validated template and improvement in biomarker estimation problems. We used two downstream tasks: cuffless blood pressure estimation as a regression problem and a noise detection model for a labeled dataset as a classification task. To do Fig. 5: Color-coded cardiac cycles based on detected noise level so, two different datasets were used in this work. The Bio-Z dataset is used to validate the improvement of biomarker estimation using the proposed framework where the data is used to accurately predict blood pressure and the MIT-BIH ECG dataset is used to illustrate high accuracy where the dataset is labeled with different noise ratios. We train and test the model for each subject individually where the data is split to 64% for training, 16% for validation and 20% for the test set.

A. Datasets

Wearable sensors are prone to generate noisy biomedical signal in certain conditions (e.g., movement or loose sensor connection). This will easily impact the downstream task for

which we use the physiological signal. Therefore, it is important to validate the data for a robust prediction. An important advantage of the proposed framework is its ability to distinguish the noisy segments in a fully unlabeled dataset using a single validated segment. However, it is equally important to demonstrate the robustness of this model in a setting where the data is labeled with specific noise metric (e.g., signal to noise ratio). Therefore, we used the two datasets listed below to demonstrate the application of the end-to-end data validator in a real-world setting.

1) Bio-Z Dataset: The first dataset used in this study comprised of Bio-impedance signal collected in an IRB-approved research study at Texas A&M University (IRB2020–0090F). Bio-impedance (Bio-Z) is an electrical signal which can be measured by injecting current to skin and capturing the voltage difference. The variation in Bio-Z corresponds to changes in blood pressure over time at the position where the sensor is placed [48]. We use the Bio-Z signals which were measured from the radial arteries. To capture a higher range of blood pressure, participants were asked to do physical activities during successive trials. Each trial contains a total of 8 minutes, 30 seconds of an initial resting period followed by 210 seconds of hand gripping exercise to elevate the blood pressure and a recovery period for 4 minutes in the end. A Finapres NOVA device was used at to measure beat-to-beat diastolic and systolic blood pressure as ground truth reference measurements. More than 150 complete 8-minute trials from 14 different participants were used for the BP estimation experiment. The data from 5 participants were removed due to large amount of noise where the DTW distance between all cluster representatives and clean template was significantly high. This shows that the data validator framework distinguished all the clusters as noisy. Therefore, the data from those participants were not included in the experiments. All the signals were downsampled to 1k Hz and the amplitude was normalized between 0 and 1 for the data validator subnetwork. However, the original amplitude range was kept intact for the BP estimation task.

2) MIT-BIH ECG Dataset: MIT-BIH Noise Dataset is comprised of 12.5 hours ECG recording data where 3.5 hours of the recording contains common noise in ambulatory ECG waveforms including baseline wander (BW), muscle artifact (MA), electrode motion artifact (EM). The electrode motion artifact source of noise is the most challenging among all noise variations as it cannot be removed by filtering, unlike other noise groups [49], [18]. The clean subset of this dataset was constructed using two samples of 118 and 119 from the MIT-BHI Arrhythmia Database [50] where noisy segments are contaminated with various ranges of noise which resulted in a range of Signal-to-Noise (SNR) ratios [–6, 24].

B. Time-series Validation

Segmented data samples were validated in the data preprocessing pipeline, where a confidence score is generated for each segment (e.g., each cardiac cycle of Bio-Z or ECG). As we can see in Fig. 5, validated cardiac cycles are correctly distinguished from the noisy segments (annotated in red) using the proposed framework. Furthermore, Fig. 4 demonstrates the color-coded Bio-Z time-series data where the blue segments demonstrated high confidence while the red segments demonstrate lower confidence. The validated data points, i.e., the points which belong to the clean cluster, are further used for biomarker

estimation (e.g., BP estimation) and noisy segments from the noisy cluster are discarded in the biomarker estimation task in order to ensure high-quality data.

C. Validated Cardiac Cycles

After the aforementioned steps, the segments are passed into the framework and only a subset of segments or more specifically, cardiac cycles with validated morphology are kept for further biomarker estimation. Fig. 7 demonstrates such a subset where the corresponding morphology is validated. The purple line illustrates the mean of all the validated cardiac cycles and the dashed lines demonstrate the range of morphology variation within one standard deviation from the average morphology in the validated cluster.

D. DTW Distance for Noisy Cluster Detection

The cluster representatives, when color-coded by the normalized DTW distance from the verified template, can be seen in Fig. 6. As we can clearly see, the dark blue cardiac cycle represents the clean cluster representative, while the red cardiac cycle represents the noisy cluster where the normalized DTW distance to the clean template is the highest. Furthermore, Fig. 8 illustrates the latent space where each point is color-coded by the normalized DTW distance to the verified cardiac cycle. As we can see the points with larger distances to the clean cluster centroid explained in the previous section, also have higher DTW distances to the verified cluster centroid.

E. GMM Clustering

Fig. 9 illustrates an example of clustering with GMMs in the latent space, where we can see the variation of morphologies of the waveform closest to the representative of each cluster. Lastly, we can see the variation in the Gaussian distributions of these clusters. As we can see, the distribution of the first cluster is significantly different compared to the other two clusters which contain relatively clean segments.

F. Noise removal quantification

As the main application of the proposed framework is to distinguish the corrupted morphologies of cardiac cycles and remove them, it is necessary to evaluate this individual component of the end-to-end framework. Fig. 7 shows an example of validated subset of cardiac cycles. As we can see, the validated cardiac cycles follow similar morphology with minimal amount of corruptness, hence a low amount of variation in the morphology of the validated cardiac cycles. As a result, in order to evaluate the result of noise removal framework, we compare the variation in morphology before and after using the noise removal framework. To do so, we compute the standard deviation of the DTW distance of cardiac cycles to the verified template in the original dataset and compare it to the distribution of the DTW distance of cardiac cycles to the verified template in the validated subset. Fig. 10 shows the impact of noise removal of the framework on the reduction DTW distance to the verified morphology. The result shows a 59.92% reduction in the standard deviation of DTW distance to the template when using the proposed framework. Moreover, for the data with relatively high noise rate (i.e., with a value more than 3 standard deviations of the mean), the noise removal module reduced the variation by 84.4%. This explains the

fact that the proposed framework is able to reduce the corrupted morphology in different settings with low and high noise rates.

G. Latent Space Dimensionality Selection

One of the most important steps in the proposed end-to-end framework is to define the number of dimensions of the latent space. In order to select the best option for the latent space dimension, we use different clustering metrics to assess the result of clustering using a different number of clusters. For example, the Silhouette score measures the similarity of points belonging to the same cluster with a range from −1 to 1. In Fig. 12, we observe the result of having different number of clusters when changing the dimensionality of the VAE latent space. It is clear that lower latent spaces result in better clustering performance which yields to better detection of noisy and clean cardiac cycles. Furthermore, Fig. 13 demonstrates the effect of changing number of clusters on the Silhouette score. It should be noted that using two other metrics, AIC (Akaike's Information Criteria) and BIC (Bayesian Information Criteria) resulted in a similar trend as shown above. As the dataset used in this work was segmented to normalized cardiac cycles, the detection of variation in their morphology can be done using lower dimensional latent spaces.

H. Latent Space Visualization Techniques

As shown in the previous section, lower dimensional latent spaces resulted in better clustering results for noisy segment detection. However, it is important to illustrate the visualization of different latent spaces with higher dimensions. As a result, we compared the result of visualizing higher dimensional latent spaces using three different methods, t-SNE (t-distributed stochastic neighbor embedding) [51], UMAP (Uniform Manifold Approximation) [52] and PCA (Principal Component Analysis). In this experiment, a VAE with an 8-dimensional latent space was trained and used in the end-to-end framework. Fig. 11 shows the result of using t-SNE, UMAP and PCA accordingly, where green points correspond to the validated cardiac cycles while the red points correspond to the noisy cardiac cycles. As we can see, the noisy segments are isolated compared to clean(green) cardiac cycles in all the plots.

I. Training Configuration

The variational autoencoder framework used in this work is implemented in PyTorch and trained using 4 GTX 1080ti GPUs using Adam optimizer with a learning rate of 10^{-3}, a weight decay factor of 10^{-5} and over 200 number of epochs.

J. Noise Classification in MIT-BIH Dataset

The MIT-BIH noise dataset was used in an unsupervised fashion to assess the accuracy of noise detection using the proposed framework. Fig. 14 demonstrates the latent space where the data is contaminated with two extreme cases, i.e., SNR = 24 where the contaminated segments were relatively less noisy (a,b) and SNR = −6, which means the contaminated segments were extremely noisy (c,d).It is obvious that where the noisy segments are closer to the clean ones in terms of morphology (SNR = 24), there is less distinction between the clusters in the latent space versus where the contaminated segments are easily

distinguishable in the case of SNR = −6. The Gaussian distributions of these two extreme cases are demonstrated in Fig. 14 (b,d). As we can clearly see, the Gaussian distribution of signals with a low SNR value is considerably different compared to the Gaussian distribution with a high SNR value. In other words, the cardiac cycles with significantly low SNR values can be discarded where the cardiac cycles with high SNR values can be kept for any further biomarker estimation relative to the generated confidence metric. Using the proposed framework, we were able to obtain 92% accuracy in detecting low SNR segments.

K. Blood Pressure Estimation Using Bio-Impedance Waveform Data

The validated subset of Bio-Z cardiac cycles was used to estimate the systolic and diastolic blood pressure. Table I summarizes the result of using validated cardiac cycles for BP estimation when using three baseline models, XGBoost, a deep learning model comprised of 3 convolutional layers and a deep learning model with two layers of convolutional layers as well as three layers of the bi-directional recurrent network. As we can see, using the proposed framework to validate signal quality resulted in significant improvement in the result of cuffless blood pressure estimation by an average of 2.67 mmHg reduction in root mean square error (RMSE) of Diastolic Blood pressure (DBP), an average of 2.13 mmHg reduction in RMSE of systolic blood pressure (SBP), an average of 0.28 increase in Pearson correlation of DBP estimation and an average of 0.08 increase in Pearson correlation of SBP estimation. Furthermore, the validator method was compared against different filtering techniques for a comprehensive comparison. As Table II shows, the proposed framework achieved the best performance compared to other filtering techniques. Fig. 15 demonstrates the Bland-Altman plots of the systolic and diastolic blood pressure estimation when the XGBoost model used the validated data by the proposed framework.

V. FUTURE DIRECTIONS

Using the proposed framework, we demonstrated a systematic method to quantify the variation between different biomedical signal morphologies to detect noisy segments without having the labeled noisy cardiac cycles in a periodic biomedical signal. However, there is a lack of understanding of the variation of feature space where subject-dependent and time-dependent morphology variations exist. Therefore, a chief direction of this work in the future is to provide tools for understanding the aforementioned variations using the mapped latent distributions. Moreover, generalizability for a fair biomarker estimation method remains a challenge. this framework can be utilized to find the right training set with overlapping feature distribution and with a high value of mutual information for adapting to an unseen subject in a generalizable setting. Another limitation of the presented work is the lack of multi-dimensional feature space with different data modalities in each training set. Thus, we plan to expand the proposed framework for multidimensional waveform data. To do so, we plan to explore having multiple validated templates in a multimodal setting where different types of noise can be eliminated. Additionally, the proposed algorithm will be explored to detect uncertainties of the ground truth biomarker output space in order to increase the confidence in label prediction, where the correlation between variation in the feature space and output space will be analyzed. By using these approaches, we aim to continue progressing toward a robust and generalizable modeling for biomarker estimation

as well as providing higher quality data for the implementation of precise algorithms, especially in the healthcare field where integration of AI-based pipelines into clinical workflows for promoting fair and equitable healthcare solutions is more challenging.

VI. Conclusion

In this work, we presented a method to detect anomalies in the biomedical waveform data using a Variational Autoencoder, Gaussian Mixture Models and minimal utilization of Dynamic Time Warping which allows for an effective quantification of the amount of data corruptness and confidence in the data usability. The proposed architecture consists of convolutional neural network layers to extract spatial information to further detect variation in waveform morphology. The model has been extensively validated on two different biomedical data modalities, bio-impedance and ECG. In conclusion, our model has the following advantages:

- Data Flexibility: The proposed method uses a fully unsupervised mechanism for the training which does not require any previous domain expert to label the data points in advance.
- Model Flexibility: The proposed distance metric can be incorporated into the loss function of any deep learning model to reduce the uncertainty in the prediction tasks and improve the metrics used to assess the outcome.
- Generalizability: The proposed framework can detect uncertainties in different biomedical data modalities across time.
- Data explainability: Using the generated distance metric, one can use the confidence metric to assess data quality and visualize various parts of time-series data to validate which input sequences to use for label prediction with high certainty.

Acknowledgments

This work was supported in part by the National Institute of Health, under grant 1R01HL151240.

References

[1]. Dimitrov DV, “Medical internet of things and big data in healthcare,” Healthcare informatics research, vol. 22, no. 3, pp. 156–163, 2016. [PubMed: 27525156]

[2]. Gao W, Emaminejad S, Nyein HYY, Challa S, Chen K, Peck A, Fahad HM, Ota H, Shiraki H, Kiriya D et al. , “Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis,” Nature, vol. 529, no. 7587, pp. 509–514, 2016. [PubMed: 26819044]

[3]. Liu W, Fang X, Chen Q, Li Y, and Li T, “Reliability analysis of an integrated device of ecg, ppg and pressure pulse wave for cardiovascular disease,” Microelectronics Reliability, vol. 87, pp. 183–187, 2018.

[4]. Ibrahim B and Jafari R, “Cuffless blood pressure monitoring from a wristband with calibration-free algorithms for sensing location based on bio-impedance sensor array and autoencoder,” Scientific reports, vol. 12, no. 1, pp. 1–14, 2022. [PubMed: 34992227]

[5]. Karraz G, “Effect of adaptive line enhancement filters on noise cancellation in ecg signals,” SJEE, vol. 18, no. 3, pp. 291–302, 2021.

[6]. Ardagna D, Cappiello C, Sama W, and Vitali M, “Context-aware datá quality assessment for big data,” Future Generation Computer Systems, vol. 89, pp. 548–562, 2018.

[7]. Kerr KA, Norris T, and Stockdale R, “The strategic management of data quality in healthcare,” Health Informatics Journal, vol. 14, no. 4, pp. 259–266, 2008. [PubMed: 19008276]

[8]. Xu Y, Zhu L, Jiang L, and Yang Y, “Faster meta update strategy for noise-robust deep learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 144–153.

[9]. Zhao Y, Chen J, and Oymak S, “On the role of dataset quality and heterogeneity in model confidence,” arXiv preprint arXiv:2002.09831, 2020.

[10]. Hsu J, Chiu W, and Yeung S, “Darcnn: Domain adaptive region-based convolutional neural network for unsupervised instance segmentation in biomedical images,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1003–1012.

[11]. van der Wal D, Jhun I, Laklouk I, Nirschl J, Richer L, Rojansky R, Theparee T, Wheeler J, Sander J, Feng F et al. , “Biological data annotation via a human-augmenting ai-based labeling system,” NPJ digital medicine, vol. 4, no. 1, pp. 1–7, 2021. [PubMed: 33398041]

[12]. Messenger JC, Ho KK, Young CH, Slattery LE, Draoui JC, Curtis JP, Dehmer GJ, Grover FL, Mirro MJ, Reynolds MR et al. , “The national cardiovascular data registry (ncdr) data quality brief: the ncdr data quality program in 2012,” Journal of the American College of Cardiology, vol. 60, no. 16, pp. 1484–1488, 2012. [PubMed: 22999725]

[13]. Shaheen MY, “Ai in healthcare: medical and socio-economic benefits and challenges,” ScienceOpen Preprints, 2021.

[14]. Frenay B and Verleysen M, “Classification in the presence of labeĺ noise: a survey,” IEEE transactions on neural networks and learning systems, vol. 25, no. 5, pp. 845–869, 2013.

[15]. Mohd Apandi ZF, Ikeura R, Hayakawa S, and Tsutsumi S, “An analysis of the effects of noisy electrocardiogram signal on heartbeat detection performance,” Bioengineering, vol. 7, no. 2, p. 53, 2020. [PubMed: 32517214]

[16]. Pereira T, Gadhoumi K, Ma M, Liu X, Xiao R, Colorado RA, Keenan KJ, Meisel K, and Hu X, “A supervised approach to robust photoplethysmography quality assessment,” IEEE journal of biomedical and health informatics, vol. 24, no. 3, pp. 649–657, 2019. [PubMed: 30951482]

[17]. Thiagarajan JJ, Rajan D, and Sattigeri P, “Understanding behavior of clinical models under domain shifts,” arXiv preprint arXiv:1809.07806, 2018.

[18]. Moody GB, Muldrow W, and Mark RG, “A noise stress test for arrhythmia detectors,” Computers in cardiology, vol. 11, no. 3, pp. 381–384, 1984.

[19]. Prashar N, Dogra J, Sood M, and Jain S, “Removal of electromyography noise from ecg for high performance biomedical systems,” Network Biology, vol. 8, no. 1, pp. 12–24, 2018.

[20]. Sayadi O and Shamsollahi MB, “Ecg denoising and compression using a modified extended kalman filter structure,” IEEE transactions on biomedical engineering, vol. 55, no. 9, pp. 2240–2248, 2008. [PubMed: 18713693]

[21]. Mohammed TA, Hassan AE, and Ferikoglu A, “Independent component analysis and extended kalman filter for ecg signal filtering,” in 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE). IEEE, 2021, pp. 1–6.

[22]. McSharry PE, Clifford GD, Tarassenko L, and Smith LA, “A dynamical model for generating synthetic electrocardiogram signals,” IEEE transactions on biomedical engineering, vol. 50, no. 3, pp. 289–294, 2003. [PubMed: 12669985]

[23]. Hesar HD and Mohebbi M, “An adaptive kalman filter bank for ecg denoising,” IEEE journal of biomedical and health informatics, vol. 25, no. 1, pp. 13–21, 2020.

[24]. Gaamouri S, Bousbia-Salah M, and Hamdi R, “Performance study of neural network unscented kalman filter for denoising ecg signal,” in International conference on the Sciences of Electronics, Technologies of Information and Telecommunications. Springer, 2018, pp. 14–23.

[25]. Hawkins DM, Identification of outliers. Springer, 1980, vol. 11.

[26]. Lane T and Brodley CE, “An application of machine learning to anomaly detection,” in Proceedings of the 20th national information systems security conference, vol. 377. Baltimore, USA, 1997, pp. 366–380.

[27]. Shaukat K, Alam TM, Luo S, Shabbir S, Hameed IA, Li J, Abbas SK, and Javed U, "A review of time-series anomaly detection techniques: A step to future perspectives," in Future of Information and Communication Conference. Springer, 2021, pp. 865–877.

[28]. Blázquez-García A, Conde A, Mori U, and Lozano JA, "A review on outlier/anomaly detection in time series data," ACM Computing Surveys (CSUR), vol. 54, no. 3, pp. 1–33, 2021.

[29]. Malhotra P, Vig L, Shroff G, Agarwal P et al. , "Long short term memory networks for anomaly detection in time series," in Proceedings, vol. 89, 2015, pp. 89–94.

[30]. Chauhan S and Vig L, "Anomaly detection in ecg time signals via deep long short-term memory networks," in 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2015, pp. 1–7.

[31]. Munir M, Siddiqui SA, Dengel A, and Ahmed S, "Deepant: A deep learning approach for unsupervised anomaly detection in time series," Ieee Access, vol. 7, pp. 1991–2005, 2018.

[32]. Deng A and Hooi B, "Graph neural network-based anomaly detection in multivariate time series," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 5, 2021, pp. 4027–4035.

[33]. Zhang A, Zhao X, and Wang L, "Cnn and lstm based encoder-decoder for anomaly detection in multivariate time series," in 2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), vol. 5. IEEE, 2021, pp. 571–575.

[34]. Thill M, Konen W, Wang H, and Bäck T, "Temporal convolutional autoencoder for unsupervised anomaly detection in time series," Applied Soft Computing, vol. 112, p. 107751, 2021.

[35]. Von Schleinitz J, Graf M, Trutschnig W, and Schroder A, "Vasp: An¨ autoencoder-based approach for multivariate anomaly detection and robust time series prediction with application in motorsport," Engineering Applications of Artificial Intelligence, vol. 104, p. 104354, 2021.

[36]. Chen R-Q, Shi G-H, Zhao W-L, and Liang C-H, "A joint model for it operation series prediction and anomaly detection," Neurocomputing, vol. 448, pp. 130–139, 2021.

[37]. Borghesi A, Bartolini A, Lombardi M, Milano M, and Benini L, "A semisupervised autoencoder-based approach for anomaly detection in high performance computing systems," Engineering Applications of Artificial Intelligence, vol. 85, pp. 634–644, 2019.

[38]. Wang W, Lyu G, Shi Y, and Liang X, "Time series clustering based on dynamic time warping," in 2018 IEEE 9th international conference on software engineering and service science (ICSESS). IEEE, 2018, pp. 487–490.

[39]. Martinez J, Sel K, Mortazavi BJ, and Jafari R, "Boosted-springdtw for comprehensive feature extraction of ppg signals," IEEE Open Journal of Engineering in Medicine and Biology, vol. 3, pp. 78–85, 2022. [PubMed: 35873901]

[40]. Guo C, Jia H, and Zhang N, "Time series clustering based on ica for stock data analysis," in 2008 4th international conference on wireless communications, networking and mobile computing. IEEE, 2008, pp. 1–4.

[41]. Sai Madiraju N, Sadat SM, Fisher D, and Karimabadi H, "Deep temporal clustering: Fully unsupervised learning of time-domain features," arXiv e-prints, pp. arXiv–1802, 2018.

[42]. Ma Q, Zheng J, Li S, and Cottrell GW, "Learning representations for time series clustering," Advances in neural information processing systems, vol. 32, 2019.

[43]. Kingma DP and Welling M, "Auto-encoding variational bayes," arXiv preprint arXiv:1312.6114, 2013.

[44]. An J and Cho S, "Variational autoencoder based anomaly detection using reconstruction probability," Special Lecture on IE, vol. 2, no. 1, pp. 1–18, 2015.

[45]. Azzalini D, Bonali L, and Amigoni F, "A minimally supervised approach based on variational autoencoders for anomaly detection in autonomous robots," IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 2985–2992, 2021.

[46]. Rousseeuw S, "A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, no. 20, p. 53.

[47]. Schwarz G, "Estimating the dimension of a model," The annals of statistics, pp. 461–464, 1978.

[48]. Ibrahim B and Jafari R, "Cuffless blood pressure monitoring from an array of wrist bio-impedance sensors using subject-specific regression models: Proof of concept," IEEE

transactions on biomedical circuits and systems, vol. 13, no. 6, pp. 1723–1735, 2019. [PubMed: 31603828]

[49]. PhysioBank P, "Physionet: components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215–e220, 2000. [PubMed: 10851218]

[50]. Moody GB and Mark RG, "The impact of the mit-bih arrhythmia database," IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45–50, 2001. [PubMed: 11446209]

[51]. Van der Maaten L and Hinton G, "Visualizing data using t-sne." Journal of machine learning research, vol. 9, no. 11, 2008.

[52]. McInnes L, Healy J, and Melville J, "Umap: Uniform manifold approximation and projection for dimension reduction," arXiv preprint arXiv:1802.03426, 2018.

Fig. 1:
Variational Autoencoder-based framework for periodic biomedical signal validation

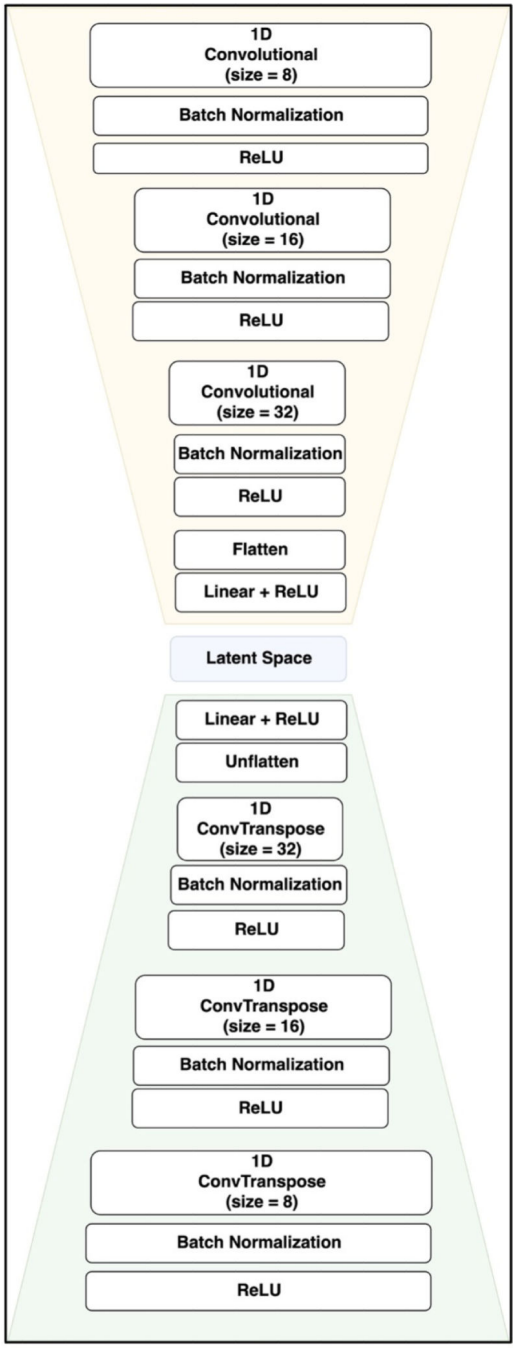


Fig. 2:
Variational Autoencoder-based framework architecture

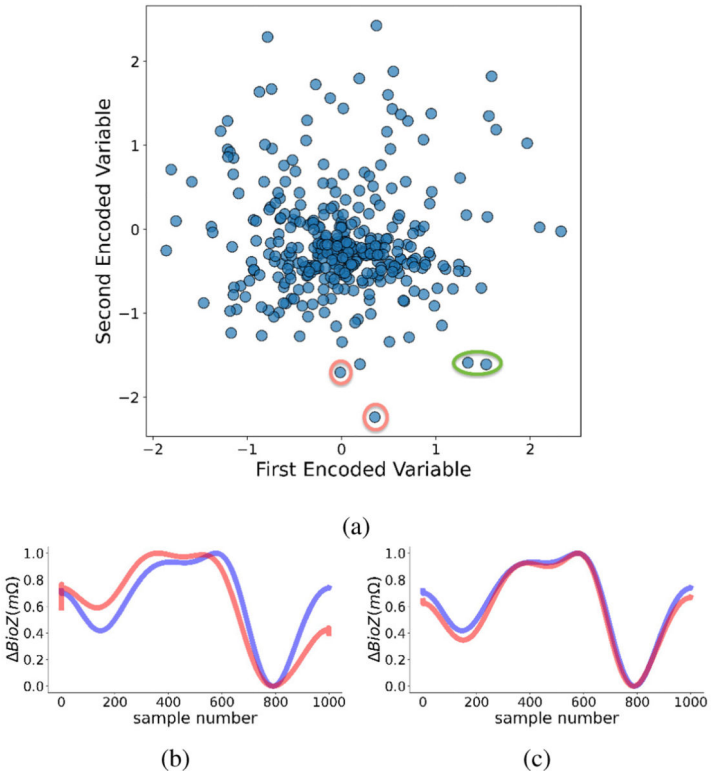


Fig. 3:

(a) Latent representation example (b) Morphologies of orange annotated points (c) Morphologies of green annotated points

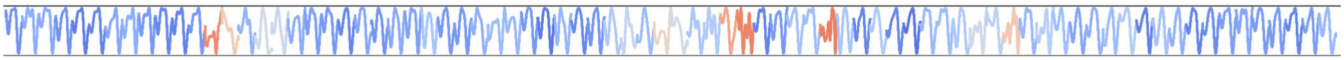


Fig. 4:
Color-coded time-series visualization based on the detected noise ratio

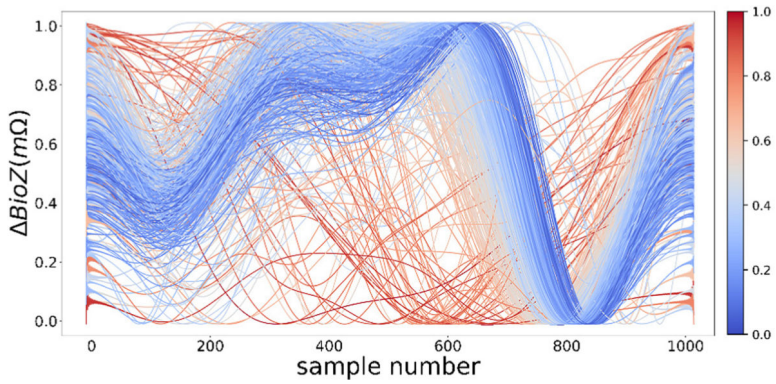


Fig. 5:
Color-coded cardiac cycles based on detected noise level

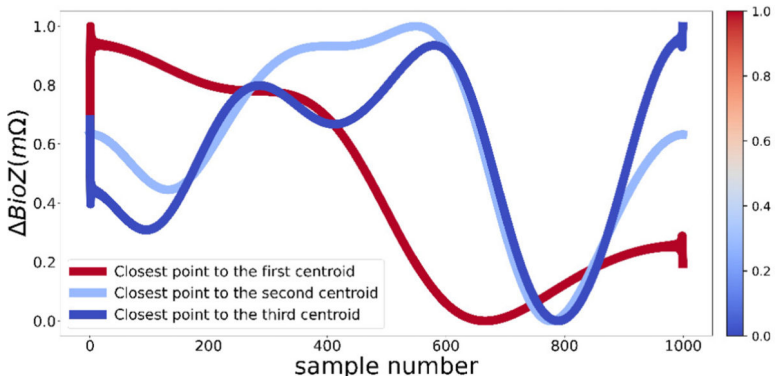


Fig. 6:

Cluster prototypes color coded by the DTW distance from a clean template. The color code demonstrates the DTW distance value where blue represents closer prototypes to the validated template and red means farther prototypes to the validated template.

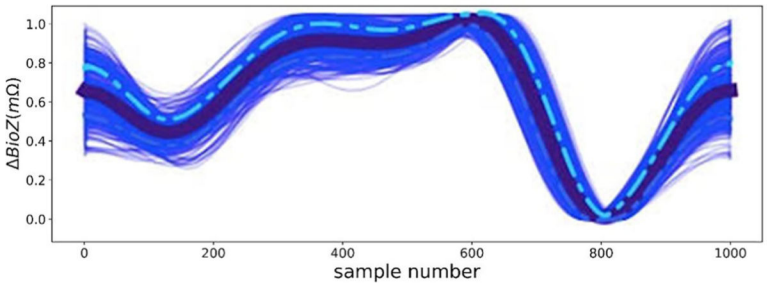


Fig. 7:

Visualization of validated cardiac cycles extracted from the cluster containing clean segments

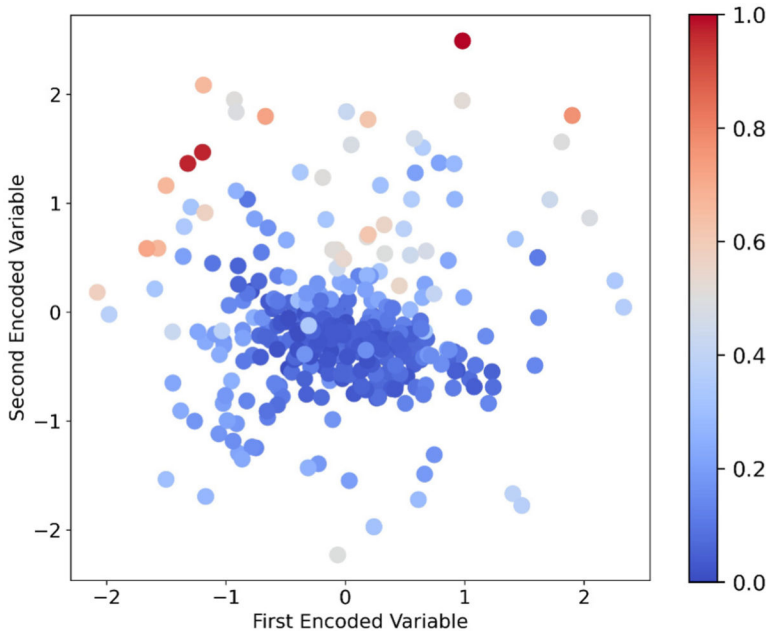
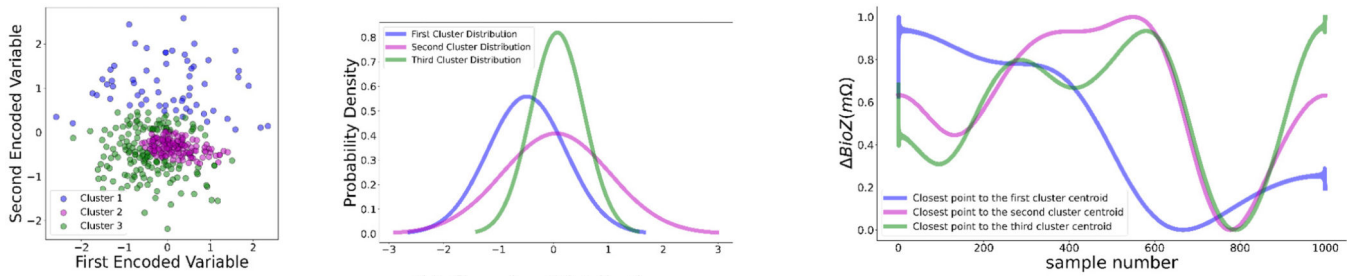


Fig. 8:

Latent space color coded by the DTW distance to the verified morphology. The color code represents the amount of noise in each segment estimated based on the DTW distance of each embedded point in the latent space to the validated template where blue means clean cardiac cycles and red means noisier cardiac cycles.



(a) Latent Space (b) Gaussian Distributions (c) Clusters Prototype

Fig. 9:

Example of GMM clustering with corresponding morphology variation across different clusters' cardiac cycle representatives and Gaussian distributions

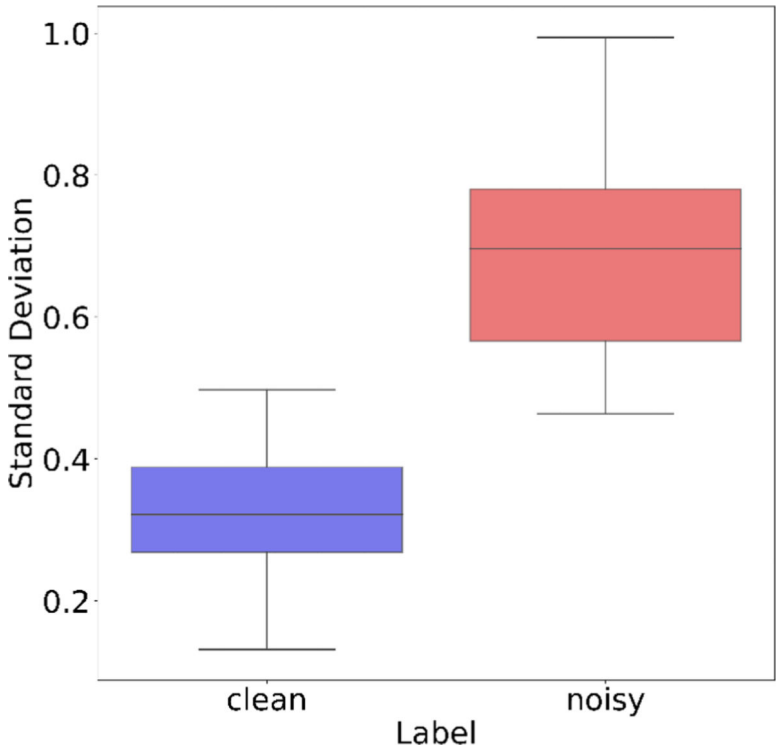
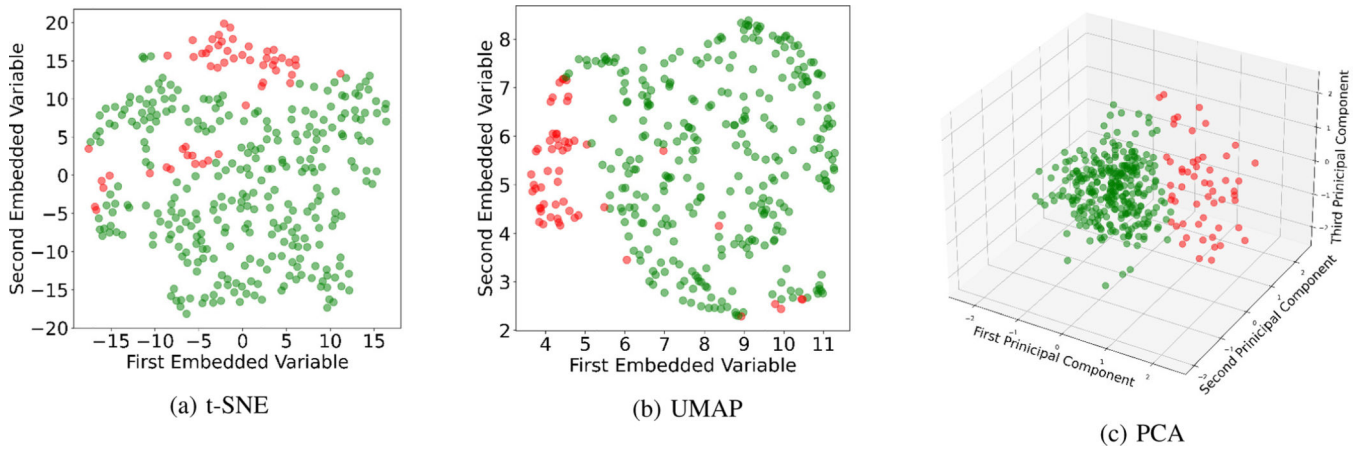


Fig. 10:

Distribution of standard deviation in DTW distance to the verified morphology. The difference in standard deviation is statistically significant ($p<0.05$).



(a) t-SNE (b) UMAP (c) PCA

Fig. 11:

Visualization comparison for high dimensional latent spaces where red points correspond to the noisy clusters and green points correspond to the clusters with validated morphology.

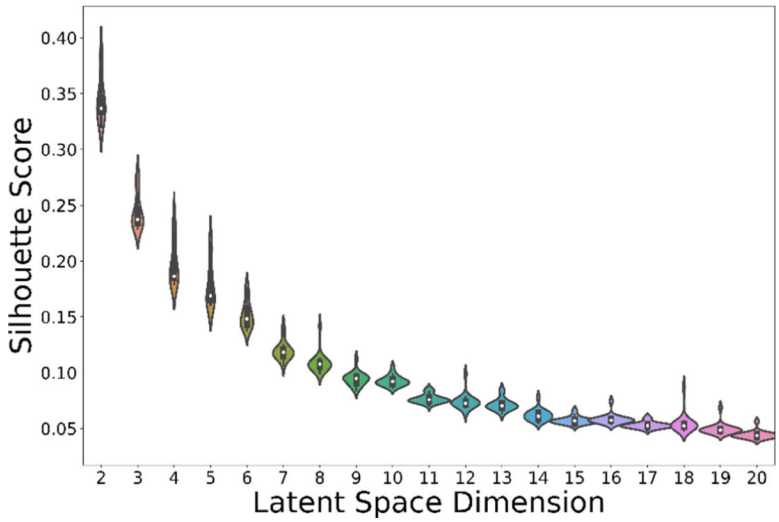
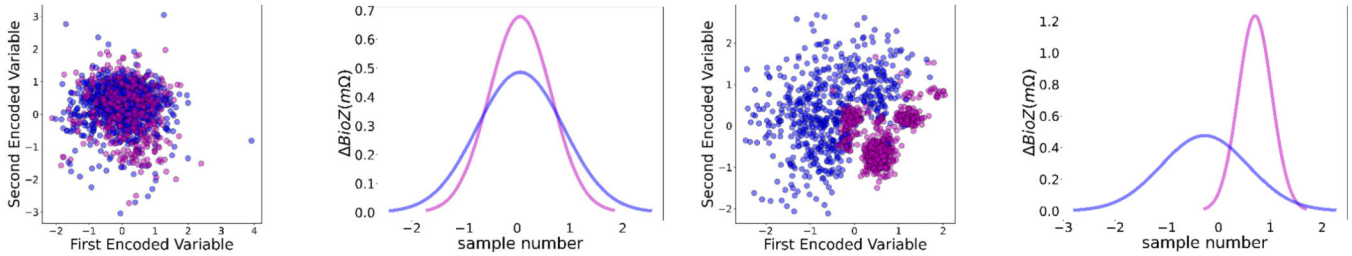


Fig. 12:
Variation of Silhouette score for clustering different latent spaces

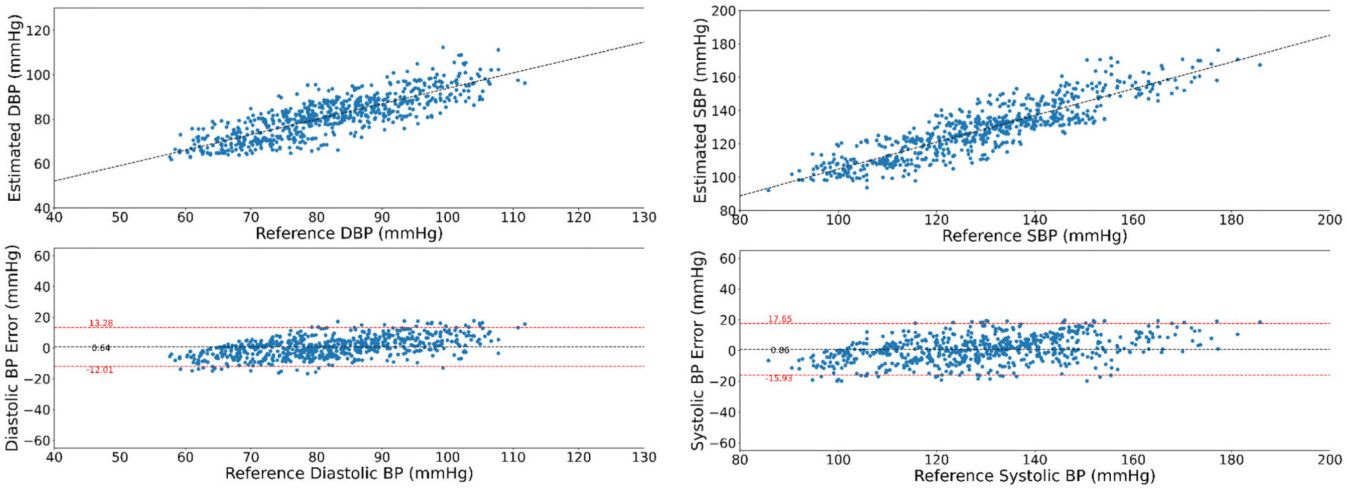
Fig. 13:
Variation of Silhouette score for latent space with different dimensions



(a) Latent space (b) Gaussian distributions (c) Latent space (d) Gaussian distributions

Fig. 14:

Latent space and Gaussian distributions when SNR = 24 (a,b) vs SNR = −6 (c,d), where purple represents the ‘clean’ distribution and blue represents the ‘noisy’.



(a) DBP (b) SBP

Fig. 15:
Bland-Altman plots for the BP prediction task

TABLE I:

Blood pressure estimation result. In order to assess if the impact of adding data validator component on the result of BP estimation is statistically significant, we conduct a paired left-tail t-test on the RMSE and a paired right-tail t-test on the Correlation values of each model with and without using validated data accordingly.

		DBP			SBP		
Data	**Model**	**RMSE(mmHg)**	**RMSE Reduction**	**r(% increase)**	**RMSE(mmHg)**	**RMSE Reduction**	**r(% increase)**
Raw Bio-Z waveform	XGBoost	9.17 ± 13.01	-	0.51	11.31 ± 11.72	-	0.74
Validated Bio-Z Waveform	**VAE + XGBoost**	**6.64 ± 1.78** *	**27.60%**	**0.80 (56.9%)** *	**8.82 ± 2.34** *	**22%**	**0.85 (14.9%)** *
Raw Bio-Z Waveform	CNN	9.21 ± 10.32	-	0.51	10.21 ± 9.64	-	0.76
Validated Bio-Z Waveform	**VAE + CNN**	**6.79 ± 1.97** *	**26.30%**	**0.78 (52.9%)** *	**8.92 ± 2.7** *	**12.60%**	**0.84 (9.5%)** *
Raw Bio-Z Waveform	CNN+Bi-GRU+Attn	9.58 ± 14.24	-	0.49	11.59 ± 12.63	-	0.79
Validated Bio-Z Waveform	**VAE + CNN+Bi-GRU+Attn**	**6.49 ± 1.94** *	**32.20%**	**0.78 (59.1%)** *	**8.87 ± 2.6** *	**23.50%**	**0.84 (6.3%)** *

*The corresponding p-value for the reduction in RMSE and improvement in correlation was < 0.05 which demonstrates the significant impact of adding the data validator component for all the models. The improved results using data validator component all met ISO standard (error < 10 mmHg).

TABLE II:

Comparison of Blood pressure estimation result when utilizing different filtering techniques against the proposed framework

		DBP		SBP	
Data	**Model**	**RMSE(mmHg)**	**r**	**RMSE(mmHg)**	**r**
Bio-Z Waveform + Band-Pass Filter	XGBoost	10.21 ± 15.97	0.35	12.27 ± 14.27	0.52
Bio-Z Waveform + Savitzky-Golay	XGBoost	9.57 ± 12.7	0.62	13.64 ± 11.32	0.54
Bio-Z Waveform + Median Filter	XGBoost	8.73 ± 7.64	0.42	10.28 ± 9.8	0.31
Validated Bio-Z Waveform	**VAE + XGBoost**	**6.64 ± 1.78**	**0.80**	**8.82 ± 2.34**	**0.85**

IEEE J Biomed Health Inform. Author manuscript; available in PMC 2025 March 28.

TABLE III:

The comparison of Signal to Noise Ratio values and standard deviation of DTW distances before and after applying the noise detection framework in the presence of different noise sources

Data	DTW SD Before Noise Removal	DTW SD After Noise Removal*	SNR Before Noise Removal	SNR After Noise Removal*
Raw Bio-Z waveform	0.8	0.3	12	23
Bio-Z Waveform + Gaussian Noise	1.2	0.2	8	18
Bio-Z Waveform + Powerline Noise	1.5	0.4	10	18
Bio-Z Waveform + White Noise	2.4	0.4	5	17

IEEE J Biomed Health Inform. Author manuscript; available in PMC 2025 March 28.