Uncertainty-Inspired Multi-Task Learning in Arbitrary Scenarios of ECG Monitoring

Xingyao Wang, Hongxiang Gao, Caiyun Ma, Tingting Zhu, Feng Yang, Chengyu Liu *Senior Member, IEEE*, and Huazhu Fu *Senior Member, IEEE*

***Abstract*—As the scenarios for electrocardiogram (ECG) monitoring become increasingly diverse, particularly with the development of wearable ECG, the influence of ambiguous factors in diagnosis has been amplified. Reliable ECG information must be extracted from abundant noises and confusing artifacts. To address this issue, we suggest an uncertainty-inspired model for beat-level diagnosis (UI-Beat). The base architecture of UI-Beat separates heartbeat localization and event diagnosis in two branches to address the problem of heterogeneous data sources. To disentangle the epistemic and aleatoric uncertainty within one stage in a deterministic neural network, we propose a new method derived from uncertainty formulation and realize it by introducing the class-biased transformation. Then the disentangled uncertainty can be utilized to screen out noise and identify ambiguous heartbeat synchronously. The results indicate that UI-Beat can significantly improve the performance of noise detection (from 91.60% to 97.50% for real-world noise detection and from 61.40% to 82.41% for real-world artifact detection). For multi-lead ECG analysis, UI-Beat is approaching the performance upper bound in heartbeat localization (only 15 false positives and 9 false negatives out of the 175,907 heartbeats in the INCART database) and achieving a significant performance improvement in heartbeat classification through uncertainty-based cross-lead fusion compared to single-lead prediction and other state-of-the-art methods (an average improvement of 14.28% for detecting heartbeats of S and 3.37% for detecting heartbeats of V). Considering the characteristic of one-stage ECG analysis within one model, it is suggested that the proposed UI-Beat has the potential to be employed as a general model for arbitrary scenarios of ECG monitoring, with the capacity to remove invalid episodes, and realize heartbeat-level diagnosis with confidence provided.**

***Index Terms*—ECG diagnosis, uncertainty disentanglement, signal quality, ambiguous data, heartbeat localization, heartbeat classification.**

I. Introduction

This work was supported by the National Natural Science Foundation of China (62171123 and 62211530112), the National Key Research and Development Program of China (2023YFC3603600). (Xingyao Wang and Hongxiang Gao contributed equally to this work.) (Corresponding author: Chengyu Liu (chengyu@seu.edu.cn).)

Xingyao Wang, Feng Yang, and Huazhu Fu are with Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore 138632 (e-mails: wang_xingyao@ihpc.a-star.edu.sg, yangf@ihpc.a-star.edu.sg, hzfu@ieee.org).

Xingyao Wang, Hongxiang Gao, Caiyun Ma, and Chengyu Liu are with School of Instrument Science and Engineering and State Key Laboratory of Digital Medical Engineering, Southeast University, Nanjing 210096, China (e-mails: xingyao@seu.edu.cn, hongxiang_seu@seu.edu.cn, caiyun@seu.edu.cn, chengyu@seu.edu.cn).

Tingting Zhu is with Department of Engineering Science, University of Oxford (e-mail: tingting.zhu@eng.ox.ac.uk).

Heartbeat identification is the first and the most critical stages for automatic electrocardiograph (ECG) analysis. Unfortunately, the performance for automatic diagnosis of wearable ECGs has reached a bottleneck that the error accumulation of heartbeat analysis will disturb the downstream tasks and is not easily corrected. The recent researches have shown that the classic methods, such as Pan-Tompkins and its variations [1], are no longer suitable for complex situations, while the deep learning (DL) approaches have brought significant improvement in wearable ECG analysis [2], [3] . Despite improving the accuracy, the DL-based methods tend to make over-confident predictions and not provide measures of confidence in their inference. Sufficient evidence is needed to help cardiologists identify ambiguous signals and beats, especially for wearable ECGs with limited lead(s) and noise contamination. As shown in Fig. 1, the anomaly in long-term single-lead ECG is not only caused by cardiac arrhythmia, but also misdirected by ambiguous heartbeats and invalid episodes contaminated by random noise. Here, an 'invalid' ECG segment refers to a segment of collected ECG data where no discernible ECG features are detected, and no continuous three or more identifiable R-waves is found. Such segments are deemed invalid due to insufficient information for reliable analysis.

The ambiguous heartbeats in ECG can be due to several reasons, including the poor quality in ECG, interference of alternating current voltage, muscle tremor, baseline drift and even pathological factors. They generally share similar morphology with QRS-complex so far as to confuse the cardiologists [4]. In practice, the so-call artifacts in wearable ECGs with insufficient multi-lead reference could be ectopic heartbeats caused by exogenous or endogenous factors. Unfortunately, both traditional and DL-based heartbeat analysis methods do not possess reliable approaches to quantify the confidence for these ambiguous waveform. This has indirectly led to a deliberate focus on improving sensitivity in the deployment of QRS-complex location module, shifting the workload to the secondary checking stage, such as extra selection of heartbeat templates. Similarly, noises are inevitable for wearable ECGs, causing extra signal quality assessment (SQA) [5], [6].

Interestingly, the two above scenarios correspond to the two types of uncertainty, respectively - epistemic and aleatoric uncertainty. By definition, epistemic uncertainty is inherent to the model, caused by a lack of training

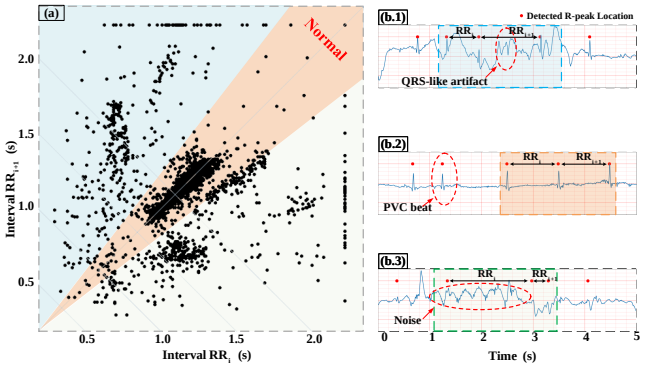


Fig. 1. (a) is the Lorenz plot for a long-term single-lead ECG of a patient with arrhythmia. Lorenz plot is a critical tool for comprehensively visualizing heart rate variability (HRV) using the plotting mode of beat-by-beat extraction of the RR intervals. The x-axis coordinate of the plot represents the RR interval before the i^{th} beat, while the y-axis coordinate represents the RR interval after. Based on the scatter in Lorenz plot, as well as the location and characteristics of the ectopic heartbeat, cardiologists may recognize the arrhythmia pattern instinctively. (b.1-3) show that the aberrant heartbeat cannot be separated from a QRS-like artifact and ambiguous noise if one solely relies on the results of the R-peak localisation. The accumulation of such confusion may hinder subsequent tasks in ECG analysis.

data and aleatoric uncertainty originates from the inherent noise in data and hence irreducible with more data [7]. Therefore, invalid signal episodes can be considered out-of-distribution (OoD) data relating to epistemic uncertainty, whereas ambiguous heartbeats can be considered to be pertaining to aleatoric uncertainty. Hence disentangling these two uncertainties and reasoning about each independently is crucial for wearable ECG analysis. The representative methods so far for uncertainty quantification are Monte-Carlo Dropout [8] and Deep Ensembles [9], which require multiple forward passes in inference. Therefore, they are unsuitable for heartbeat analysis in long-term ECGs, which should be a lightweight first step in ECG analysis.

Except for the interference from noise and ambiguous signal, another bottleneck in DL-based diagnostic for ECGs is the lack of and inconsistency in annotations. In-hospital cardiovascular examination predominantly rely on shot-term, 12-lead resting ECGs. However, the annotation becomes exceedingly challenging when it extends to dynamic monitoring scenarios, especially with non-standard ECG signals. The difficulty in annotation first leads to heterogeneity in data distribution across different ECG analysis tasks. For example, compared to tasks like heartbeat classification and P wave localization, the identification of R-peaks in ECG is relatively straightforward, resulting in more comprehensive databases containing R-peak annotations. On the other hand, databases related to abnormal heartbeat or atrial anomaly detection are exceedingly scarce and often have incomplete annotations. Furthermore, the blurring boundaries between different categories results in severe inconsistencies in annotations, even from the same expert at different times [10].

To address the aforementioned issues, we present an uncertainty-inspired framework for beat-level analysis of arbitrary ECGs, including a multi-task learning module to address the heterogeneity between heartbeat localization and classification, an uncertainty disentanglement method to estimate epistemic and aleatoric uncertainty for noise detection and ambiguous heartbeat identification. The intuition of our work is to accomplish the necessary modules for heartbeat analysis within one stage, so that substantial costs will not be required when executing noise screening out, heartbeat localization and heartbeat classification in a step-wise manner. For convenience, we named our uncertainty-inspired heartbeat analysis method as UI-Beat. The main contributions of this paper are summarized as follows:

1) For the first time we integrate uncertainty estimation into the beat-wise analysis of ECG within one model, and achieve the identification of noise and ambiguous heartbeat at the same time based on a new uncertainty disentanglement method while ensuring the high precision in heartbeat localization and classification.

2) In order to address the issues of label incomplete and task heterogeneity, we develop a cross-source learning method to facilitate information interaction between heartbeat localization task and event detection task.

3) To address the inadequacy of current detection method for artifact noise and ambiguous heartbeats, we construct the first database for ambiguous ECG signals. All the annotated ambiguous ECG episodes are published as AM-ECG database in https://github.com/mato00/SEU-AM-ECG.

4) We comprehensively validate the performance of the proposed framework on noise detection, ambiguous heartbeat identification, heartbeat localization and classification. According to the experimental results, it is indicated that UI-Beat can provide clinically trustworthy diagnostic conclusions, and be a potential unified model for multi-modal cardiovascular signal analysis.

II. Related Works

A. ECG Segmentation and Classification

So far, the majority of intelligent ECG analysis have been divided into two main areas: ECG segmentation and classification. ECG segmentation includes heartbeat localization based on the detection of R-peaks [11], ECG delineation containing QRS-complex, P-waves and T-waves [12], and ECG feature generation and clustering derived from subwave detection [13]. The significance of segmentation lies in its ability to provide physiological indicators for long-term ECG analysis. Current research on ECG classification mainly focuses on end-to-end supervised learning [14], including the design of NNs tailored for the extraction of pseudo-periodic features in ECGs [15], and studies on label distribution learning for multi-label issues for ECG classification [16], [17]. Considering the variance in different ECG collection approaches, self-supervised learning has been explored through pretraining on integrated ECG databases. Experiments have shown that pretrained AI models can enhance classification performance for downstream tasks, especially

for rare anomalies [18], [19]. Due to the inconsistent distribution of annotations for ECG segmentation and diagnosis, most studies still approach ECG analysis as a multi-step problem [20], [21]. However, ECG semantic delineation and diagnosis should be interdependent. Recently, some research has considered how to integrate ECG segmentation and classification tasks within a single model. For instance, Gao *et al.* adopted continual learning to implement ECG segmentation and abnormality detection within a backbone network [22]. Moreover, a significant bottleneck in ECG analysis is to establish trustworthy AI models for clinic. In fact, for diagnosis of 12-lead resting ECGs, there has been exploratory research on the impact of uncertainty estimation in AI while deploying AI in clinic [23]. For fine-grained beat-level ECG analysis, the uncertainty inherent in heartbeat variants will cause the unstable diagnostic results, and contribute to inconsistent data distributions and annotations.

B. Multi-Source Multi-Task Learning

Multi-task learning aims to learn multiple tasks simultaneously through the interdependence between different tasks [24]. In multi-task learning, a prevalent approach involves utilizing shared hidden layers within a single model to address various tasks, known as "hard-sharing". Despite its popularity, this technique often leads to diminished performance, especially in multi-source scenario [25], [26]. The key issue in multi-source learning is the deficiency of shared samples among tasks induced by data heterogeneity. For multi-task learning, the approach to address multi-source problems can be summarized as establishing mechanisms for cross-task information transfer [24], and evaluating the relationships among tasks is the most direct strategy. For example, operating under the premise that tasks facing the same sample are not mutually independent, [27] proposes a method to enhance the learning of cross-task consistency. Zhou *et al.* [28] measures the similarity between tasks based on adversarial learning and utilize the information from similar tasks to improve the performance of independent tasks. Beyond directly measuring inter-task correlations, knowledge distillation and transfer are also employed to align representations of single-tasks with those of multi-task models, thereby mitigating training imbalances caused by task heterogeneity [29], [30]. In recent years, researchers have also explored modeling the inter-task correlations based on the theory of multivariate Gaussian process, which has led to the design of training procedure for neural networks (NNs) [31], [32]. In the context of ECG diagnostics, tasks with simpler annotations typically have a more comprehensive distribution, yet there is a lack of annotations for more challenging tasks.

C. Uncertainty in Deterministic Models

The majority of uncertainty estimation approaches in deep learning are under Bayesian framework, that is to assume the posterior distribution of model parameters given training data [7]–[9]. However, these methods are either unable to scale to large databases and model architectures, suffer from low-quality uncertainty measurement, or require expensive ensemble. Quantifying uncertainty in deterministic neural networks is more applicable since they own the lowest computational cost and not modify the inference procedure. Entropy measurement of predictive results is first attempted to estimate uncertainty in deterministic models [33]. Unfortunately, entropy measurement is often not indicative of OoD data [34]. Feature space modeling offers a different approach [35]. The essence is to establish a constraint in the latent distribution induced by iD and OoD data, so that once a previously unseen point with high distance (low density) added to the training data, the feature space should be revealed with low distance (high density) [36]. Several related studies have already existed, like Kong et al. [37] explicitly confuses the feature space referring to Brownian motion using two independent Encoders, and Amersfoort et al. [35] utilizes Radial Basis Function (RBF) network to exclude OoD data straying in feature-space.

III. Methods

A. ECG Analysis Framework

The proposed model for ECG heartbeat analysis encompasses two distinct tasks: heartbeat localization (R-peak localization) and heartbeat diagnosis. For heartbeat localization, the proposed method will first detects the coarse QRS-complex interval for each heartbeat, then identifies the R-peak within that interval. The nature of these two tasks is distinct: localization is to detect QRS-complex in ECG, while diagnosis requires an integrated decision considering the contextual information. Moreover, due to the difference in annotation difficulty, any publicly available ECG database that contains annotations for R-peaks can be utilized for training in heartbeat localization task, and such databases tend to own a more comprehensive ECG distribution. In contrast, ECG databases with annotations for abnormal heartbeats are rarer and mostly constructed by clean ECGs.

Consequently, as illustrated in Fig. 2, in order to achieve the beat-level diagnosis, the proposed UI-Beat model employs structurally identical yet independent encoders and decoders for the tasks of heartbeat and event detection, respectively. In ECG, an event such as sinus rhythm (SR) and ventricular arrhythmia (VA) in Fig. 2, or atrial arrhythmia (AA), may contain multiple heartbeats. Ultimately, heartbeat diagnosis integrates the predicated probability sequence for heartbeat localization and event detection, that is, assigning a diagnostic type to each localized heartbeat within an event interval. Another aim of this design is to do cross learning for the two tasks asynchronously owing to the heterogeneous data distribution in heartbeat localization and classification. The details will be demonstrated in the next section.

While synchronously localizing and diagnosing each heartbeat, frame-level uncertainty estimation is conducted based on a new proposed uncertainty disentanglement formulation. Here we utilize $\mathbb{U}$ represents the predictive uncertainty, and $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$ the epistemic and aleatoric uncer-

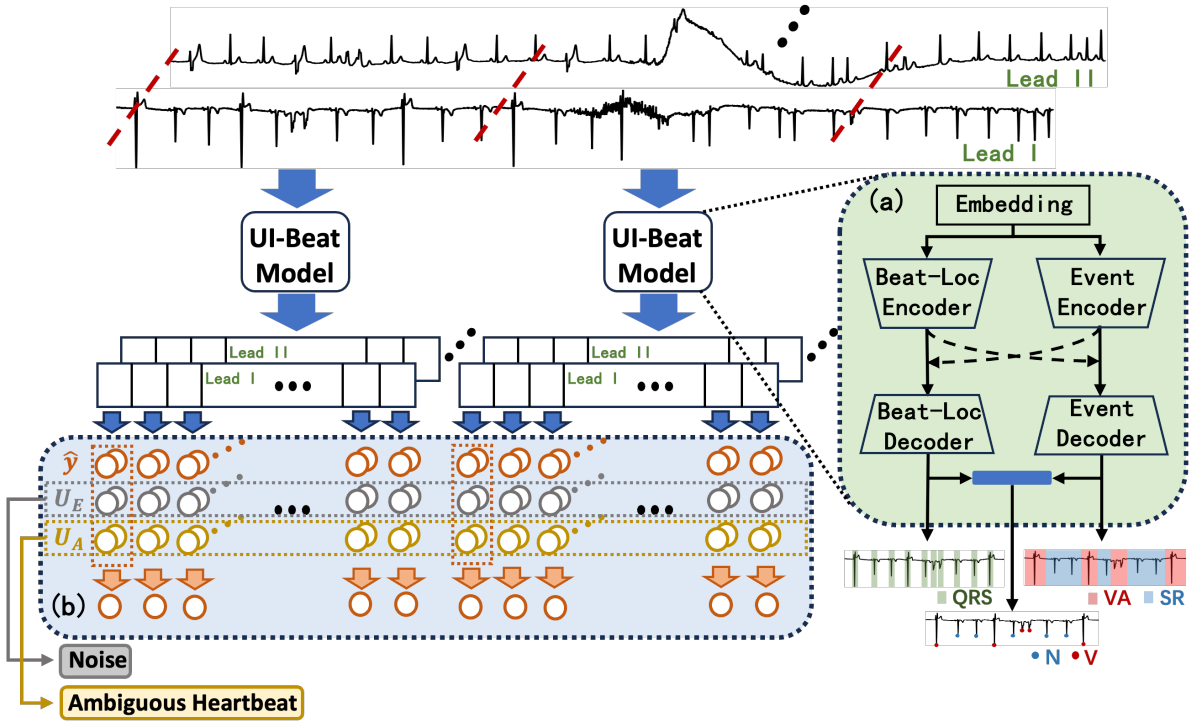


Fig. 2. Overall ECG analysis pipeline based on the proposed UI-Beat model. An arbitrary ECG with any number of leads is segmented into fragments for input to the UI-Beat model at first. (a) illustrates the basic architecture of UI-Beat model designed for the multi-task and multi-source scenario in this work. The results of heartbeat classification are obtained by merging the outputs of heartbeat localization and event detection. In (b), based on the proposed uncertainty disentanglement method, UI-Beat provides frame-level outputs including $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$. The outputs from the lead with the highest confidence at each frame is selected as the ultimate output for that frame of the ECG. Meanwhile, $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$ are respectively responsible for detecting noise and ambiguous heartbeat.

tainty, respectively. In UI-Beat, the decoupled $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$ respectively provide the capabilities for noise filtration and the detection of ambiguous heartbeats. Meanwhile, through uncertainty-based cross-lead fusion, the output from the lead with the highest confidence is adopted as the final prediction, and UI-Beat will ultimately realize noise detection, ambiguous beat identification, heartbeat localization and diagnosis within one stage.

B. Cross-Source Learning

As aforementioned, since the heartbeat localization task and heartbeat classification task are heterogeneous, we consider the heartbeat localization database as $\mathcal{D}_l = \{X_l, Y_l\}$, the heartbeat classification database as $\mathcal{D}_c = \{X_c, Y_c\}$, and the event database $\mathcal{D}_e = \{X_e, Y_e\}$. The event database is constructed from $\mathcal{D}_c$ as $\mathcal{D}_e = \{X_e, Y_e\}$, where X_e contains intervals, and Y_e denotes labels associated with these intervals. While $\mathcal{D}_e$ was derived from $\mathcal{D}_c$ by grouping sequences of abnormal beats, it is also possible to derive $\mathcal{D}_c$ from $\mathcal{D}_e$. In this alternative process, each heartbeat within an abnormal interval in $\mathcal{D}_e$ would be assigned a label reflecting the specific abnormal rhythm type, thereby creating a heartbeat-level classification dataset. Accordingly, we have revised the description to accurately reflect the relationship between $\mathcal{D}_c$ and $\mathcal{D}_e$. The encoders and decoders of the two tasks are defined as f_l and f_e, g_l and g_e, respectively. Thus, for heartbeat localization task, we have $\mathbf{z}_l = f_l(x_l); \hat{y}_l = g_l(\mathbf{z}_l)$, and for event detection task, $\mathbf{z}_e = f_e(x_e); \hat{y}_e = g_e(\mathbf{z}_e)$. Before entering the encoders, we uniformly performed embedding on the input ECG episode. The reason for employing event detection as a precursor task to heartbeat classification lies in the further compression of input ECG signals in latent space. Since ECGs are non-stationary signals prone to noise interference, more compression can mitigate the impact of noise on the output.

In order to induce the utilization of contextual information in heartbeat localization within this multi-task, multi-source learning framework, and to introduce a broader data distribution of heartbeat for event detection, we design and implement the inter-task interaction. This interaction is realized by transforming and aligning $f_e(x_l)$ with z_l and then adding them when input x_l is provided. A similar approach is conducted with input x_e. While training, the supervision from $\mathcal{D}_l$ is exclusively linked to the parameters updates for f_l and g_l, and the supervision from $\mathcal{D}_e$ is solely associated with the parameters of f_e and g_e. However, for the heartbeat classification task with database D_c, since it is an intersection between heartbeat localization and event detection, we merge $g_l(f_l(x_c))$ with $g_e(f_e(x_c))$. Given that the output for heartbeat localization is a probability sequence for binary states (QRS-complex and non-QRS interval), we treat it as an attention tensor and utilize a channel-wise multiplication

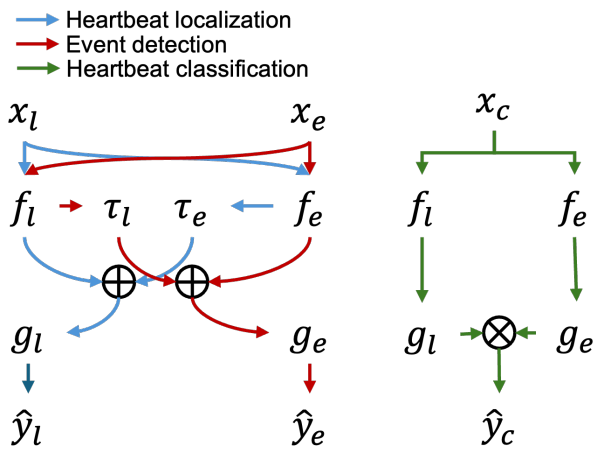


Fig. 3. Illustration of the data flow during model's forward propagation in heartbeat localization, event detection, and heartbeat classification.

method to integrate it with the output from event detection. The formulas for cross-task information interaction and the probability output for heartbeat classification are as follows:

$$\mathbf{z}_l^{'} = \mathbf{z}_l + \tau_e(f_e(x_l)), \tag{1}$$

$$\mathbf{z}_e^{'} = \mathbf{z}_e + \tau_l(f_l(x_e)), \tag{2}$$

$$\hat{y}_c = g_l(f_l(x_c)) \cdot g_e(f_e(x_c)). \tag{3}$$

To clarify cross-source learning in our proposed method, we have provided the data flow of the model's forward propagation in the flowchart (Fig. 3).

For convenience, the parameters for f_l and f_e are defined as θ_l and θ_e, while for decoders are ω_l and ω_e. In the following sections, the parameters of the combined encoder and decoder will be indicated as θ and ω. During training, the optimization objective function in our scenario can be defined as:

$$\begin{aligned}\min_{\theta,\omega} \mathcal{L}(\theta,\omega,\mathcal{D}) =& \ell_{\text{loc}}(\theta_l,\omega_l,(X_l,Y_l)) \\ &+ \ell_{\text{evn}}(\theta_e,\omega_e,(X_e,Y_e)) \\ &+ \ell_{\text{cls}}(\theta,\omega,(X_c,Y_c)),\end{aligned} \tag{4}$$

where ℓ_{loc}, ℓ_{evn}, and ℓ_{cls} are the loss functions for heartbeat localization, event detection, and heartbeat classification respectively. For UI-Beat training, we simply utilize cross entropy for the loss function of each task.

C. Frame-wise Uncertainty Disentanglement

According to [38], given a data point $\{x, y\} \in \mathbb{D}$ and the prediction $\hat{y}$ for x, the predictive uncertainty $\mathbb{H}\left[E_{P(\theta|\mathcal{D})}[p(Y \mid x, \theta)]\right]$ upper-bounds the epistemic uncertainty, where epistemic uncertainty is quantified as the mutual information $\mathbb{I}[Y; \theta \mid x, \mathcal{D}]$ brought by θ on $\hat{y}$. The formulation is defined as follows:

$$\underbrace{\mathbb{I}[Y, \theta \mid x, \mathcal{D}]}_{\text{epistemic}} = \underbrace{\mathbb{H}[Y \mid x, \mathcal{D}]}_{\text{predictive}} - \underbrace{E_{P(\theta|\mathcal{D})}[\mathbb{H}[Y \mid x, \theta]]}_{\text{aleatoric (for iD } x)}. \tag{5}$$

The $\mathbb{U}_{\mathbb{E}} = \mathbb{I}[Y; \theta \mid x, \mathcal{D}]$ can be isolated from $\mathbb{U}_{\mathbb{A}}$ by sampling θ from $P(\theta \mid \mathcal{D})$. This approach estimates $\mathbb{U}_{\mathbb{A}}$ as the expected entropy $E_{P(\theta|\mathcal{D})}[\mathbb{H}[Y \mid x, \theta]]$, allowing $\mathbb{U}_{\mathbb{E}}$ to capture the uncertainty due to variability in θ itself. Therefore, the way is infeasible for a deterministic network with a solely-determined θ. Here we assume that $\mathbb{U}_{\mathbb{E}}$ can be regarded as the information gain brought from the predictive label of x to the encoded representations, $\mathbf{z} = f_\theta(x)$. That is to say, if to identify a new data input x^* as label y^*, how much more evidence can the model capture from the data?

As a sequence-to-sequence task, the predicted result on each frame $\hat{y}_\tau$ is considered to be determined by the representations of this frame on the decision layer. Therefore, the inference procedure obeys the Markov chain $x \to \mathbf{z} \to \mathbf{z}_\tau \to \hat{y}_\tau$. Thus $\mathbf{z} = f_\theta(x)$ and $\hat{y}_\tau = g_\omega(\mathbf{z}_\tau)$, then we have $\mathbb{I}$ between the $\mathbf{z}$ and $\hat{y}$ as follows:

$$\begin{aligned}\mathbb{I}[\hat{y}, \theta \mid x, \mathcal{D}] &= \mathbb{I}[\hat{y}, \mathbf{z} \mid \mathcal{D}] \\ &= \sum_{\tau=1}^{T} \mathbb{I}[\hat{y}_\tau; \mathbf{z}_\tau \mid \mathcal{D}] \\ &= \sum_{\tau=1}^{T} (\mathbb{H}[\mathbf{z}_\tau \mid \mathcal{D}] - \mathbb{H}[\mathbf{z}_\tau \mid \hat{y}_\tau, \mathcal{D}]) \\ &= \sum_{\tau=1}^{T} D_{\text{KL}}\left(\underbrace{P(\mathbf{z}_\tau \mid \hat{y}_\tau, \mathcal{D})}_{\text{class-biased}} \parallel P(\mathbf{z}_\tau \mid \mathcal{D})\right),\end{aligned} \tag{6}$$

where T is the total number of the frames.

As Eqn.6 shows, $\mathbb{U}_{\mathbb{E}}$ can be deemed as the sum of the KL divergence between the two distributions of $\mathbf{z}_\tau$, the class-biased distribution and the ground truth distribution, which is learned from the training set. We expect a feasible solution to estimate $P(\mathbf{z}_\tau \mid \hat{y}_\tau, \mathcal{D})$ by producing discrepancy for the posterior distribution belonging to a certain class $c \in [1, K]$, $p(\mathbf{z}_\tau \mid \hat{y}_\tau^c, \mathcal{D})$. Thus we introduce ϕ to transform $\mathbf{z}_\tau$ into the class-biased representation:

$$\begin{aligned}P(\mathbf{z}_\tau \mid \hat{y}_\tau, \mathcal{D}) &= \sum_c P(\mathbf{z}_\tau, \phi^c \mid \hat{y}_\tau^c, \mathcal{D}) \\ &= \sum_c P(\mathbf{z}_\tau \mid \hat{y}_\tau^c, \phi^c, \mathcal{D}) P(\phi^c \mid \hat{y}_\tau^c, \mathcal{D}) \\ &= \frac{1}{K} \sum_c P(\mathbf{z}_\tau \mid \hat{y}_\tau^c, \phi^c, \mathcal{D}),\end{aligned} \tag{7}$$

Here the ϕ^c is responsible for finding the evidence of the target class as far as possible, and the prior distribution, $p(\phi^c \mid \hat{y}_\tau^c, \mathcal{D})$ is set as the constant $1/K$, which is under the assumption that the evidence distribution of the target class c is well determined by the model after learning from $\mathcal{D}$.

Recall and update the aforementioned Markov chain, and we can have $x \xrightarrow{f_\theta} \mathbf{z} \xrightarrow{\phi} \mathbf{z}_\tau \xrightarrow{g_\omega} \hat{y}_\tau$. In a general sense, this Markov chain can be utilized for image under pixel-level modeling, and time series data under frame-level modeling. Thus we have

$$P(\mathbf{z}_\tau \mid \hat{y}_\tau, \phi^c, \mathcal{D}) = P(\mathbf{z}_\tau \mid \hat{y}_\tau^c, \phi^c(\mathbf{z})). \tag{8}$$

In our thoughts, the transformation ϕ^c should be class-biased and preferentially capture features for the target class

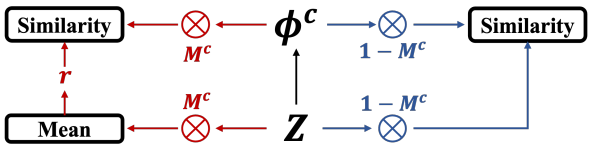


Fig. 4. Scheme of training ϕ^c.

on each frame. For iD data, $\phi^c(\mathbf{z})$ should be consistent with $\mathbf{z}$ since the evidence distribution is well-determined, so that $\mathbb{U}_{\mathbb{E}}$ will be minimal. Conversely, we expect $\phi^c(\mathbf{z})$ to be inconsistent with $\mathbf{z}$ when facing OoD data, inducing an increasing $\mathbb{U}_{\mathbb{E}}$. Defining the frame-level representation after class-biased transformation is $\mathbf{z}_\tau^c$, the $\mathbb{U}_{\mathbb{E}}$ of each frame can be derived from Eqn. 6 as:

$$\begin{aligned}\mathbb{U}_{\mathbb{E}} &= \sum_{c=0}^{K-1} D_{\mathrm{KL}}(P(\mathbf{z}_\tau \mid \hat{y}_\tau^c, \phi^c(\mathbf{z}))\|P(\mathbf{z}_\tau \mid \mathbf{z})) \\ &= \sum_{c=0}^{K-1} g(\mathbf{z}_\tau^c)\log\frac{g(\mathbf{z}_\tau^c)}{g(\mathbf{z}_\tau)}.\end{aligned} \tag{9}$$

For estimation of $\mathbb{U}_{\mathbb{A}}$, we utilize the expected entropy of $\hat{y}$, which is a straightforward approach referring to [38]. The formulation is shown as follows:

$$\mathbb{U}_{\mathbb{A}} = -\frac{1}{K}\sum_{c=1}^{K-1} E[\log g(z_\tau)]. \tag{10}$$

D. Class-biased Transformation

After establishing the formulation for uncertainty estimation, the next question is how to obtain the aforementioned class-biased transformation ϕ. Since ϕ^c should maintain the feature invariance of the target class, we need to train ϕ^c to generate representations approaching the marginal distribution of the class, which is defined as $\mathbf{r}_\tau^c \sim P(\mathbf{z}_\tau \mid y_\tau = c, \mathcal{D})$. Thus, we have the objective function for ϕ^c as follows:

$$\begin{aligned}\mathcal{L}_{\phi^c} = \frac{1}{T}\sum_{\tau=0}^{T-1}\{&d\left(\mathbf{z}_\tau^c, \mathbf{r}_\tau^c \mid y_\tau = c\right) \\ &+ d\left(\mathbf{z}_\tau^c, \mathbf{z}_\tau \mid y_\tau \neq c\right)\},\end{aligned} \tag{11}$$

Here the cosine distance is chosen here to measure the distance between representations, which is:

$$d\left(\mathbf{z}_1, \mathbf{z}_2\right) = -\frac{\mathbf{z}_1}{\|\mathbf{z}_1\|_2}\cdot\frac{\mathbf{z}_2}{\|\mathbf{z}_2\|_2}, \tag{12}$$

where $\|\cdot\|_2$ denotes l_2-norm.

To estimate $\mathbf{r}^\mathbf{c}$, we adopt the Monte-Carlo approximation, which is shown as follows:

$$\mathbf{r}^\mathbf{c} \approx \frac{1}{N}\sum_{i=0}^{N-1}\overline{\mathbf{z}_i \otimes M^c}, \tag{13}$$

where the M^c is responsible for excluding frame-level representations not belonging to class c. The illustration of the training scheme of ϕ^c is shown in Fig. 4. The average representation of all frames belonging to the target class c will be calculated on each sample batch. In the pursuit of an unbiased estimation of r^c, the average representations of all the training steps will be updated when the training epoch is ended. Assuming total sample size is N and batch size is B, the number of training steps is set as $M = N/B$. Given that r^c is solely determined by $P(\mathbf{z} \mid \theta, \mathcal{D})$, it is imperative to segregate the gradient updating of θ and ϕ^c while training. Otherwise, $\phi(\mathbf{z})$ and $\mathbf{z}$ will converge towards homogeneity, rendering ϕ^c incapable of effectuating a class-biased transformation on $\mathbf{z}$. The training procedure of ϕ^c is provided in Alg. 1.

Algorithm 1 Training procedure of ϕ^c

Require: M: number of training steps of one epoch
Require: α, β: step size hyperparameters
Randomly initialize θ, ϕ^c
while training **do**
  for $i \leftarrow 1$ to M **do**
    Input sample batch
    Evaluate $\nabla_\theta \mathcal{L}\left(f_\theta\right)$
    Update $\theta \leftarrow \theta - \alpha\nabla_\theta\mathcal{L}\left(f_\theta\right)$
    Calculate $\mathbf{r}_i^c$
  end for
  $r^c = \frac{1}{M}\sum_i r_i^c$
  Update $\phi^c \leftarrow \phi^c - \beta\nabla_{\phi^c}\mathcal{L}_{\phi^c}$
end while

E. Uncertainty-based Cross-lead Fusion

Algorithm 2 Inference procedure with cross-lead fusion

Require: L and T: number of lead(s) and frames
Require: T_e and T_a: window size for screening noises and identifying ambiguous heartbeats
Require: V_e and V_a: threshold for $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$
Predict the probability sequence $\hat{y} = \{\hat{y}_\tau \mid \tau = 0, \cdots, T-1\}$
Calculate $\{\mathbb{U}_{\mathbb{E}\tau} \mid \tau = 0, \cdots, T-1\}$ and $\{\mathbb{U}_{\mathbb{A}\tau} \mid \tau = 0, \cdots, T-1\}$ from Eqn. 9 and Eqn. 10
$\mathbb{U}_{\mathbb{E}} \leftarrow \frac{1}{T_e}\sum_{\tau=0}^{T_e-1}\mathbb{U}_{\mathbb{E}\tau}$
Reject the output from the lead with $\mathbb{U}_{\mathbb{E}} > V_e$
$\mathbb{U}_\tau \leftarrow \mathbb{U}_{\mathbb{E}\tau} + \mathbb{U}_{\mathbb{A}\tau}$
for $\tau \leftarrow 1$ to T_e **do**
  $\hat{y}_\tau \leftarrow \hat{y}_\tau^l, \mathbb{U}_{\mathbb{A}\tau} \leftarrow \mathbb{U}_{\mathbb{A}\tau}^{\ l}$ s. t. $l = \arg\min\{\mathbb{U}_\tau^l\}_{l=1}^L$
end for
Calculate heartbeat locations $\{\mathbf{R}_i\}_{i=1}^S$ from $\hat{y}$
for $i \leftarrow 1$ to S **do**
  $\mathbb{U}_{\mathbb{A}} \leftarrow \frac{1}{T_a}\sum_{\tau=R_i-T_a/2}^{\tau=R_i+T_a/2}\mathbb{U}_{\mathbb{A}\tau}$
  Accept the certain heartbeats with $\mathbb{U}_{\mathbb{A}} \leq V_a$ and ambiguous ones with $\mathbb{U}_{\mathbb{A}} > V_a$
end for

The main benefit of disentangling epistemic and aleatoric uncertainty is to distinguish OoD and iD ambiguous data. This is not trivial since an iD sample ought to cause a low

epistemic uncertainty regardless of whether it is ambiguous (with high aleatoric uncertainty) or not [38]. Therefore, the aleatoric uncertainty would be arbitrary for a noise episode, and on this account the ambiguous heartbeats and invalid noises are confusing when facing overall uncertainty quantification. We should firstly screen out the invalid noises according to the estimated $\mathbb{U}_{\mathbb{E}}$ and then identify the ambiguous heartbeats within the effective ECG episodes based on $\mathbb{U}_{\mathbb{A}}$. Here the thresholds for $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$ are defined as V_e and V_a, which will be set empirically according to the statistical distributions of $\mathbb{U}_{\mathbb{E}}$ on ECG and noise signals, and $\mathbb{U}_{\mathbb{A}}$ on certain heartbeats and ambiguous heartbeats in the subsequent sections. In Alg. 2, a scheme for cross-lead information fusion based on the proposed uncertainty estimation is provided, which is to select the output from the optimal lead with the lowest uncertainty frame-by-frame. The minimum window size is set as T_e frames for screening out invalid noises with $\mathbb{U}_{\mathbb{E}}$, and T_a frames centered on each predicted beat location for identifying ambiguous heartbeats. Since the width of the QRS-complex is typically between 60 and 100 ms, the average aleatoric uncertainty of the time window with 100 ms centered on each R-peak is utilized in this work to determine whether or not the heartbeat is ambiguous.

IV. Experimental Results

As a unified framework for wearable ECG analysis inspired by uncertainty, UI-Beat should not only own the ability to locate heartbeats and detect abnormal ones, but also to screen out noises and identify ambiguous beats within one model. Synchronously, the disentangled $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$ should alert the ambiguous heartbeats as well as their possible categories of abnormalities and corresponding evidence. In this section, we comprehensively conduct different experiments to investigate the feasibility and performance of the proposed method for heartbeat analysis in wearable ECGs.

A. Noise & Ambiguous Beat Detection

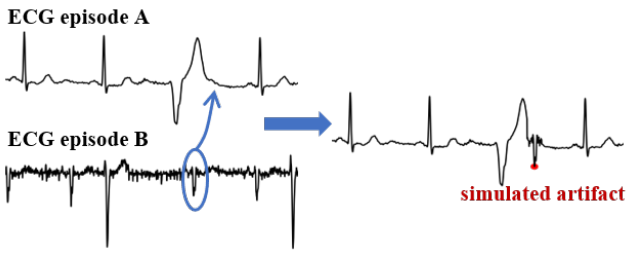
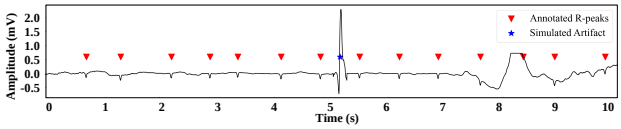
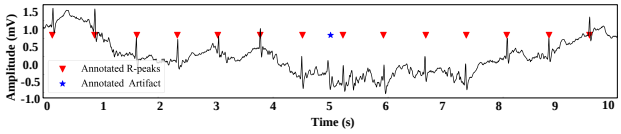


Fig. 5. Simulation of ambiguous heartbeats.

1) Database: For noise screening, 3,000 invalid 10-second episodes (without any distinguishable heartbeat) were extracted from a real-world wearable ECG database [39]. In addition, to comprehensively illustrate the advantage of the proposed uncertainty estimation method for noise detection,



(a) Instance of an ECG episode with simulated artifact



(b) Instance of the annotated artifact in real world

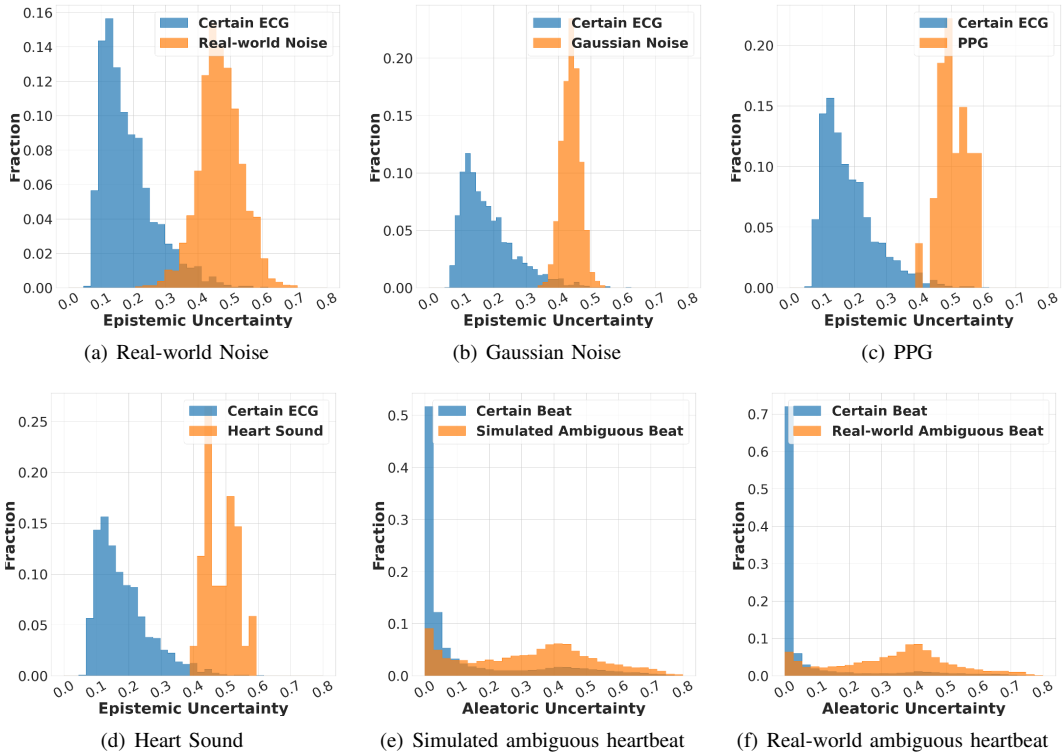
Fig. 6. (a) and (b) display the simulated and real-world instances of the ECG episodes with artifacts, respectively.

we also included two external test sets using different types of non-ECG signals, which are heart sounds and photoplethysmography (PPG) signals, and utilized CPSC2019 database (10-second ECG episodes with QRS-complex annotated) [40] as the iD effective ECG set.

For identifying ambiguous heartbeats, two data sets are constructed, including 2,000 ECG episodes with a simulated ambiguous heartbeat, which is a randomly-inserted QRS-complex from different patients and a real-world data set containing 960 samples with 1,018 manually annotated QRS-like artifact heartbeats. To ensure the correctness of recognizing heartbeats in ECGs, we often need to refer to contextual information since the absolute morphology of the artifacts (derived from noise) is close to or even equal to the QRS-complex in most cases. Therefore, the simulation will ensure a portion of ECG episodes contain an ambiguous heartbeat that has general morphological characteristics while not conforming to the contextual pattern. The visual description of generating the simulated ambiguous heartbeats is shown in Fig. 5.

To capture the ambiguous heartbeats among real-world wearable ECGs, we first looked through the records for the candidate ambiguous beats outside the morphology and rhythm patterns of normal beats (N), atrial (A), and ventricular (V) premature beats. Two physiologists then checked and annotated the candidate ambiguous heartbeats. In the end, we left the ambiguous beats on which the two physiologists provided inconsistent annotations. Besides the annotations of ambiguous beats, we also annotated certain heartbeats to evaluate the specificity of UI-Beat recognizing ambiguous heartbeats. Figs. 7(a) and 7(b) list the classic instances of the simulated and manually annotated ambiguous heartbeats. For convenience, we name the simulated and real-world databases for ambiguous heartbeats as SIM-AM-ECG and RW-AM-ECG, respectively.

2) Experimental Setup: For the evaluation of detecting ambiguous heartbeats, the positive predictive value (PPV), sensitivity (Se), and F_1 score are computed over the cumulated true positive (TP), false positive (FP) and false



(a) Real-world Noise (b) Gaussian Noise (c) PPG

(d) Heart Sound (e) Simulated ambiguous heartbeat (f) Real-world ambiguous heartbeat

Fig. 7. Statistical histograms of $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$. (a) $\sim$ (d): Comparison of the distribution of $\mathbb{U}_{\mathbb{E}}$ for UI-Beat in the presence of effective ECG and non-ECG episodes (real-world noise, Gaussian noise, PPG, and heart sound). (e) $\sim$ (f): Comparison of the distribution of $\mathbb{U}_{\mathbb{A}}$ for UI-Beat in the presence of simulated and real-world ambiguous beats.

negative (FN). And for noise screening, accuracy (Acc) is utilized as an evaluation index for classifying effective ECG and non-ECG episodes. True negative (TN) is additionally utilized in the Acc calculation. These metrics are defined as follows:

$$Se = \frac{TP}{TP + FN} \times 100\%, \tag{14}$$

$$PPV = \frac{TP}{TP + FP} \times 100\%, \tag{15}$$

$$F_1 = \frac{2 \times Se \times P_+}{Se + P_+} \times 100\%, \tag{16}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%. \tag{17}$$

We compare the noise detection results with two representative SQA methods, GbSQI [41] and kSQI [42]. GbSQI combines the features with consensus heartbeat detection. Unlike GbSQI, kSQI is developed based on the distribution characteristics of the candidate ECG episode. These two methods represent the two major directions for SQA.

3) Results: We visually compare the value distributions of the estimated $\mathbb{U}_{\mathbb{E}}$ from UI-Beat according to Eqn. 9 when facing ECG and non-ECG signals (Fig. 7). The window size T_e for screening out noise is set to 10 seconds. We can see that the estimated $\mathbb{U}_{\mathbb{E}}$ is capable of separating certain ECGs from non-ECG noises properly. The threshold V_e in Alg. 1 is set at 0.35, referring to the junction point of the two distributions. Table I shows the classification results of UI-Beat using epistemic uncertainty. Apparently, detection of non-ECG signals based on $\mathbb{U}_{\mathbb{E}}$ is significantly superior to the current SQA methods. Especially for testing on real-world noises, the proposed method achieves a performance increase of nearly 6% over GbSQI. For heart sounds and PPGs, the other two types of pseudo-periodic signals similar to ECGs, kSQI is incapable of distinguishing between them. And while maintaining the performance on detecting non-ECG episodes, UI-Beat owns over 10% performance gain over the general SQA methods in identifying certain ECGs in CPSC2019-TEST.

Similar to the experiments on noise screening, we compare the value distributions between $\mathbb{U}_{\mathbb{A}}$ on certain and ambiguous heartbeats. From Fig. 7, we can see that the estimated $\mathbb{U}_{\mathbb{A}}$ can separate ambiguous beats from certain beats. According to the junction point of the two distributions, we set the threshold V_a in Alg. 1 to 0.1 and the heartbeats with $\mathbb{U}_{\mathbb{A}}$ exceeding 0.1 will be accepted as possible artifacts. Table II shows the detection results of ambiguous heartbeats in the simulated and real-world test sets. Due to the subjectivity of the annotation, the model does not necessarily respond within the intervals of the annotated artifacts, and we exclude the samples with no response at annotated artifacts according to $\hat{y}$. According to Table II, the results of Se on SIM-AM-QRS (75.36%) and on RW-AM-QRS (88.01%) for recognizing ambiguous

TABLE I
THE PERFORMANCE COMPARISON BETWEEN THE PROPOSED UI-BEAT AND THE TWO REPRESENTATIVE SQA METHODS FOR SCREENING NON-ECG SIGNALS BASED ON EPISTEMIC UNCERTAINTY.

Method	ECG Set			Non-ECG Set								
	CPSC2019-TEST			Heart Sounds			PPGs			Real-world Noise		
	TP	FN	Acc (%)	TP	FN	Acc (%)	TP	FN	Acc (%)	TP	FN	Acc (%)
GbSQI	2411	821	74.60	1011	1307	43.62	3715	35	99.07	2748	252	91.60
kSQI	2690	542	83.23	32	2286	1.38	2031	1719	54.16	1558	1442	51.93
UI-Beat	**3033**	**99**	**96.84**	**2257**	**61**	**97.37**	**3750**	**0**	**100.00**	**2925**	**75**	**97.50**

TABLE II
THE PERFORMANCE OF DETECTING QRS-LIKE ARTIFACTS BASED ON ALEATORIC UNCERTAINTY.

Database	Certain Beats						Ambiguous Beats					
	TP	FP	FN	$Se(\%)$	$PPV(\%)$	$F_1(\%)$	TP	FP	FN	$Se(\%)$	$PPV(\%)$	$F_1(\%)$
SIM-AM-QRS	29031	57	453	98.46	99.80	99.13	1193	1110	390	75.36	51.80	61.40
RW-AM-QRS	12981	33	130	99.01	99.75	99.38	499	145	68	88.01	77.48	82.41

heartbeats indicate that $\mathbb{U}_{\mathbb{A}}$ can be utilized as an auxiliary confidence for the outputs from UI-Beat. However, there is an obvious overlap between the ambiguous and certain beats, and such confusion brings down PPV for ambiguous beats. The confounding distributions are either embodied in the contextual disturbance or the noisy annotations.

B. Heartbeat Localization for Multi-lead ECGs

1) Database: So far, the key element for heartbeat localization in multi-lead ECGs lies in how to integrate cross-lead information well. In this experiment, we use the CPSC2019-Train database [40] for training and five other independent databases with annotations of R-peaks, including CPSC2019-Test for testing the baseline performance on the single-lead heartbeat localization, and four other databases to evaluate the performance of heartbeat localization on multi-lead ECGs, including MIT-BIH Arrhythmia Database (MIT-BIH) [43], QT Database (QTDB) [44], INCART 12-lead Arrhythmia Database (INCART) [45] and LUDB 12-lead Database [46].

2) Experimental Setup: The specific implementation is shown in Alg. 1 and a 2-second overlap is adopted for long-term ECGs. Also, we compare UI-Beat against the rule-based R-peak detection method, Pan-Tompkins (P&T) [47], and several classic uncertainty-aware methods, including MC-Dropout [8] and Deep Ensemble [9], which we use an ensemble of five networks. To ensure an objective evaluation, we conduct the SQA method, GbSQI [41], as the preprocession for P&T in a multi-lead setting. All the methods that take uncertainty into account are trained with a multi-branch convolutional neural network (MBCNN) [48], which is one of the optimal methods so far in R-peak detection. For each frame, the estimated uncertainty is given, and the output sequence is made up of the frame-level likelihood that belongs to the lead with the most confidence.

Besides the general evaluation indicators (*Se*, *PPV* and F_1) mentioned above, we complement the performance loss (%) in *FN* and *FP* of a method $\mathbf{M}$ compared with the method with the minimum *FN* and *FP* for a more direct display of performance discrepancy, which is defined as follows:

$$\mathrm{FN}_{LOSS}(\mathbf{M}) = \frac{\mathrm{FN}(\mathbf{M}) - \mathrm{FN}_{min}}{\mathrm{FN}(\mathbf{M})} \times 100\%, \tag{18}$$

$$\mathrm{FP}_{LOSS}(\mathbf{M}) = \frac{\mathrm{FP}(\mathbf{M}) - \mathrm{FP}_{min}}{\mathrm{FP}(\mathbf{M})} \times 100\%. \tag{19}$$

3) Results: Table III gives a quantitative evaluation of heartbeat localization in multi-lead ECGs with uncertainty-based decision. The results over CPSC2019-TEST indicate that the baseline performance of the three uncertainty-aware QRS-complex detection methods is equivalent and far superior than P&T, of which Deep Ensemble is slightly better. According to the results over multi-lead ECG databases, it is clear that UI-Beat achieved a significant performance gap when compared with MC-Dropout and is competitive with Deep Ensemble. Furthermore, for MIT-BIH, UI-Beat reduces the FPs and FNs on heartbeat localization by more than 20% and 52%, respectively, and for INCART, the database of Holter ECGs, FPs and FNs are reduced by more than 50% and 40%, even 96.82% FPs are reduced when compare with MC-Dropout. With the aid of the proposed uncertainty estimation method, the heartbeat localization performance of UI-Beat has been close to the limit on the long-term multi-lead ECG databases. For MIT-BIH, the F_1 score achieves 99.95%, which has exceeded the limits of the current R-peak detection methods (99.86% in [49], 99.83% in [50] and 99.91% in [48]) for this database. Similarly, the cumulative errors, FPs and FNs, are only 15 and 9 for UI-Beat evaluated in INCART, approximating the limit of the performance of R-peak detection for this database.

C. Abnormal Heartbeat Classification

1) Database: For heartbeat classification, we use the MIT-BIH database [43] to construct the training and test set with different types of heartbeats. In agreement with the American Association of Medical Instrumentation (AAMI) [51] recommended practice, the four recordings containing

TABLE III
COMPARISON OF QRS-COMPLEX DETECTION PERFORMANCE ON MULTI-LEAD ECGS. EXCEPT FOR UI-BEAT, ONE RULE-BASED METHOD (P&T) AND TWO UNCERTAINTY-AWARE METHODS (MC-DROPOUT AND DEEP ENSEMBLE) ARE EVALUATED THROUGH THE ECG DATABASES.

Database	Lead Type	Method	TP	FP	$FP_{LOSS}(\%)$	FN	$FP_{LOSS}(\%)$	$Se(\%)$	$PPV(\%)$	$F_1(\%)$
CPSC2019-Test	1 Lead	P&T	36799	4599	95.11	6533	95.53	84.93	88.89	86.86
		MC-Dropout	43040	334	32.63	**292**	**0.00**	**99.33**	99.23	99.28
		5-Ensemble	43022	**225**	**0.00**	310	5.81	99.28	**99.48**	**99.38**
		UI-Beat	43024	310	27.42	308	5.19	99.29	99.28	99.29
LUDB	12 Leads	P&T+GbSQI	1763	30	53.33	22	95.45	98.77	98.33	98.55
		MC-Dropout	1784	16	12.50	1	0.00	99.94	99.11	99.53
		5-Ensemble	1784	14	0.00	1	0.00	99.94	99.22	99.58
		UI-Beat	1784	**14**	**0.00**	**1**	**0.00**	**99.94**	**99.22**	**99.58**
QTDB	2 Leads	P&T+GbSQI	69989	881	99.21	571	98.60	99.19	98.76	98.97
		MC-Dropout	70551	10	20.00	9	11.11	99.99	99.99	99.99
		5-Ensemble	70549	8	12.50	11	27.27	99.98	99.99	99.99
		UI-Beat	70552	**7**	**0.00**	**8**	**0.00**	**99.99**	**99.99**	**99.99**
MIT-BIH	2 Leads	P&T+GbSQI	101469	5828	99.59	7965	98.71	92.72	94.57	93.64
		MC-Dropout	108967	137	82.48	467	79.87	99.57	99.87	99.72
		5-Ensemble	109238	30	20.00	196	52.04	99.82	99.97	99.90
		UI-Beat	109340	**24**	**0.00**	**94**	**0.00**	**99.91**	**99.98**	**99.95**
INCART	12 Leads	P&T+GbSQI	172099	3207	99.53	3708	99.76	97.89	98.17	98.03
		MC-Dropout	175780	472	96.82	27	66.67	99.98	99.73	99.86
		5-Ensemble	175792	30	50.00	15	40.00	99.99	99.98	99.99
		UI-Beat	175798	**15**	**0.00**	**9**	**0.00**	**99.99**	**99.99**	**99.99**

TABLE IV
THE MAPPING BETWEEN AAMI AND MIT-BIH CLASSES.

Category	Class
N	Normal beat (N) Left and right bundle branch block beat (L, R) Atrial escape beat (e) Nodal (junctional) escape beat (j)
S	Atrial premature beat (A) Aberrated atrial premature beat (a) Nodal (junctional) premature beat (J) Supraventricular premature or ectopic beat (S)
V	Premature ventricular contraction (V) Ventricular escape beat (E)

paced beats are removed from the analysis, and five types of heartbeats are concluded. The details of the five heartbeat classes are described in Table IV-C1. Since N, S, and V are the three main types of heartbeats in wearable ECG analysis with relatively comprehensive annotations among open-access ECG databases, the experiments are conducted on the three categories.

Referring to the scheme introduced by De Chazal *et al* [52], MIT-BIH is divided into two datasets: the training set (DS1) and the test set (DS2). This dataset construction scheme is designed to separate the heartbeat data according to the different patients and consider balancing the classes. The detailed information on the heartbeat distribution and the recording division can be found in [52]. Except for MIT-BIH, another 12-lead ECG database with annotated heartbeats of N, S, and V, INCART [45], is utilized in our experiments, which contains 153,676 N, 1,960 S, and 20,013 V.

2) Experimental Setup: For heartbeat classification, we continue to use Se, PPV, and $F1$ score as evaluation metrics. To assess the performance of the proposed UI-Beat model in terms of uncertainty estimation, similar to the multi-lead heartbeat localization task, we evaluated the model's performance using a cross-lead fusion method while synchronously screening out noise and identifying ambiguous heartbeats, as described in Alg. 2. If the uncertainty-based cross-lead fusion approach significantly improves performance compared to single-lead recognition, it indicates that the uncertainty estimation method proposed in UI-Beat is effective. Furthermore, we also compared the UI-Beat model with current state-of-the-art (SOTA) methods in heartbeat classification, which are mainly divided into two types: methods with known and unknown heartbeat positions, where most belong to the methods with known heartbeat position. Methods with known heartbeat positions include the symbolic representation learning proposed by Niu *et al.* [21], the Pyramid-like model proposed by He *et al.* [53], and the domain-adaptive heartbeat classification method proposed by Wang *et al.* [54]. Methods with unknown heartbeat positions include the end-to-end model proposed by Yang *et al.* [55], and the deep learning-based heartbeat classification method following ECG segmentation proposed by Mathews *et al.* [56]. The UI-Beat model incorporates the heartbeat localization task during training but does not use heartbeat positions as a prior while inference. Instead, it combines the results of heartbeat localization to classify heartbeat types, thus UI-Beat belongs to the category

of methods with unknown heartbeat positions. DS1 derived from MIT-BIH database was utilized as the only training set.

3) Results: Table IV-C3 shows the comprehensive performance comparison for heartbeat classification, which contains two main parts, comparison with SOTA methods and comparison between scenarios of single-lead and uncertainty-based multi-lead fusion. According to Table IV-C3, the UI-Beat model's performance in identifying N, S, and V types of heartbeats exceeds the current SOTA, including methods that incorporate known heartbeat location information. It shows the most considerable improvement in recognizing heartbeats of S. On the DS2 derived from MIT-BIH, the $F1$ score increased by over 10% compared to the runner-up, and on the INCART database, it saw an increase of more than 15%. Besides the UI-Beat model's enhanced performance in identifying N, S, and V, the effectiveness of the proposed uncertainty estimation is evident from the comparison results between single-lead and multi-lead scenarios. For 2-lead MIT-BIH and 12-lead INCART, compared to single-lead ECG diagnosis, the uncertainty-based cross-lead fusion (Alg. 2) achieved widespread performance improvements. Specifically, for recognizing heartbeats of S, the $F1$ score improved by 4.50% and 2.39%, respectively, and for heartbeats of V, it increased by 1.52% and 3.17%, respectively.

To more directly demonstrate the role of uncertainty disentanglement and estimation in UI-Beat, we randomly selected an episode of single-lead ECG with unstable signal quality from a wearable ECG database for testing. We then visualized the diagnostic and confidence level results output by UI-Beat. The visualization results is shown in Fig. 8. In Fig. 8, the confidence (Conf.) is represented by $1 - \mathbb{U}_{\mathbb{A}}$, and the noise is screened out based on $\mathbb{U}_{\mathbb{E}}$. According to Fig. 8, UI-Beat is capable of effectively eliminating noise while making reliable heartbeat localization and diagnosis, and it offers confidence levels for heartbeats with ambiguous characteristics among different categories.

V. Discussion

For the first time we introduce uncertainty-inspired learning into heartbeat-level ECG analysis with a deterministic neural network. We develop the uncertainty paradigm into distance measurement between class-biased representations and disentangle the epistemic and aleatoric uncertainty without ensemble learning. The proposed method is named as UI-Beat. According to the definition of uncertainty, the epistemic and aleatoric uncertainty can be utilized to identify OoD data (non-ECG noise) and iD ambiguous data (ambiguous heartbeats).The experimental results suggest that UI-Beat possesses the capability of realizing noises screening, artifactambiguous heartbeat identification, high-precision heartbeat localization and heartbeat diagnosis within one-stage inference. Moreover, the proposed class-biased transformation for uncertainty disentanglement in UI-Beat is easily implemented and plug-and-play for arbitrary Encoders, which is consistent with a rapid and lightweight deployment for the initial stage of ECG analysis.

A. Noise Screening

There are two main types of noise in ECG monitoring, non-ECG signals and QRS-like artifacts. The capability of depressing the influence brought by different noises is vital in ECG analysis, especially for wearable ECG monitoring due to intense motion. In this work, we compare two classic SQA methods, kSQI and GbSQI, which represent two technical routes of SQA that extract waveform distribution characteristics and incorporate R-peak information. According to Table I, kSQI is evidently invalid for non-ECG pseudo-periodic signals, such as heart sounds and PPGs. This indicates that the SQA methods based on waveform distribution will be disturbed by periodic noises like breath and muscle tremors. For GbSQI, the generalized version of bSQI, ten QRS detectors are utilized to inform the intelligibility of QRS-complex. This idea is similar to deep ensemble in essence yet the performance is constrained by the R-peak detector. Therefore, applying uncertainty estimation in AI model for ECG diagnosis to filter out non-ECG episodes is more sensible from a performance and efficiency standpoint.

For QRS-like artifact identification, the limited research concentrates on heartbeat clustering. In 12-lead ECG monitoring, the artifacts therein can be restricted due to the multi-lead reference. However, for single-lead ECGs, the subsequent ECG analysis will be severely obstructed with plenty of suspected premature beats and long RR intervals if the frequent artifacts are not appropriately handled. According to Table II, the artifacts appear to inherit the ambiguous characteristics, as evidenced by a low PPV when recognizing ambiguous heartbeats. The identified QRS-like artifacts determined by the disentangled aleatoric uncertainty from UI-Beat and its derived indicators can be retained as auxiliaries in the final diagnostic report. By undertaking such actions, UI-Beat can significantly alleviate the verification workload caused by artifacts and ensure the reliability of subsequent ECG diagnosis tasks.

B. Heartbeat Diagnosis

The DL-based QRS-complex detection and heartbeat classification has been proved to exceed the traditional methods, especially for dynamic ECG. The unification of ECGs acquired in variational monitoring scenarios, that is to localize and identify heartbeats in any lead, is essential for improving the efficiency of ECG analysis. Traditional approaches often employ average lead and heartbeat to integrate ECG signals from different leads, which proves remarkably efficacious for multi-lead scenario. However, such methods invariably encounter limitations inherent to the performance of the algorithm. In addition, the cost for changing amount of input channels is relatively high in DL-based method. In our work, on the basis of the proposed uncertainty estimation approach, we have successfully linked the heartbeat-level diagnosis on single- and multi-lead ECGs. In fact, UI-Beat can convert single- or multi-lead, short- or long-term ECGs into batches for parallel inference and its efficiency is comparable to that

TABLE V
COMPARISON OF HEARTBEAT CLASSIFICATION PERFORMANCE ON DS2 DERIVED FROM MIT-BIH AND INCART.

Database	Lead Number	Method	N			S			V		
			Se	PPV	F1	Se	PPV	F1	Se	PPV	F1
MIT-BIH (DS2)	2	Niu *et al.* (Lead A)* [21]	98.90	97.40	98.10	76.50	76.60	76.60	85.70	94.10	89.70
		He *et al.* (Lead A)* [53]	92.00	99.00	95.37	91.00	35.00	50.56	89.00	81.00	84.81
		Wang *et al.* (Lead A)* [54]	96.06	98.31	97.17	71.92	82.89	77.02	92.01	80.16	85.68
		Yang *et al.* (Lead A) [55]	98.91	95.48	97.17	19.22	48.82	27.58	87.01	95.01	90.84
		Mathews *et al.* (Lead A) [56]	–	–	–	88.39	33.63	48.72	77.74	69.20	73.22
		UI-Beat(Lead A)	99.00	98.08	98.54	89.71	81.95	85.65	86.27	94.91	90.38
		UI-Beat	99.52	98.26	**98.88**	90.42	89.88	**90.15**	90.18	93.68	**91.90**
INCART	12	He *et al.* (Lead II)* [53]	90.30	99.30	94.59	79.40	15.40	25.80	87.00	72.70	79.21
		Wang *et al.* (Lead II)* [54]	–	–	–	44.07	37.38	40.45	77.02	92.08	83.88
		Yang *et al.* (Lead II) [55]	97.35	97.60	97.48	66.11	71.67	68.77	79.63	92.66	85.65
		UI-Beat(Lead II)	99.08	97.46	98.26	82.95	80.69	81.80	80.56	97.31	88.15
		UI-Beat	99.42	97.61	**98.50**	79.53	89.44	**84.19**	84.58	99.22	**91.32**

[1] * represents the method with heartbeat location known.

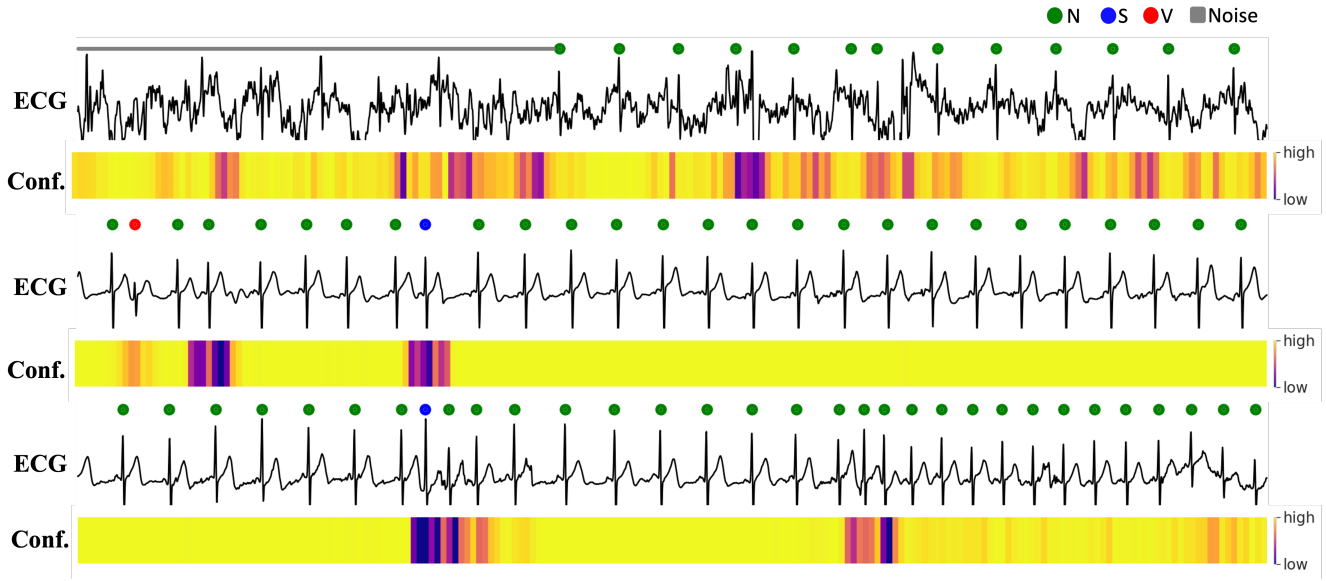


Fig. 8. Visualization of heartbeat localization and diagnosis on a wearable single-lead ECG using UI-Beat. In addition to displaying the results of heartbeat localization and classification, the figure also presents the judgement of noise and confidence level based on the disentangled $\mathbb{U}_{\mathbb{E}}$ and $\mathbb{U}_{\mathbb{A}}$. The gray line indicates the noise interval, and according to Alg. 2, all diagnostic results within this interval are rejected. The frame-level confidence is calculated by $1 - \mathbb{U}_{\mathbb{A}}$ and transformed into heat maps to intuitively showcase the correspondence among ECG signals, diagnostic outcomes, confidence levels. It is observable that when encountering noise interference or ambiguous heartbeats, the confidence is generally low, whereas for clean heartbeats with distinct features, the confidence is notably high.

of traditional methods. Notably, for multi-lead ECGs, UI-Beat can effectively integrate the cross-lead information and make decisions based on the frame-level outputs from the most confident lead.

Certainly, several areas need to be completed for UI-Beat in the future. First of all, UI-Beat focuses on class-biased transformation for producing discrepant latent representations in a deterministic neural network. As the number of abnormal types requiring diagnosis increases, the operation becomes more redundancy. In the future, in order to reduce the cost of learning, it is necessary to investigate unified learning paradigms or integrable estimation of marginal distributions for class-biased transformation. In addition, one of the most important applications of uncertainty estimation is in the field of active learning, where it is used to strengthen the model by including data with high epistemic uncertainty but low aleatoric uncertainty into the training set. Therefore, we will continue to the proposed UI-Beat in utilizing uncertainty to realize increment learning of heartbeat variations. Last but no least, since the efficiency of UI-Beat has been validated on ECG analysis in arbitrary scenarios, the proposed frame-level uncertainty estimation and uncertainty-based cross-lead fusion has the potential to be extended to multi-modal cardiovascular diagnostics.

VI. Conclusion

This paper proposes a unified heartbeat-level diagnosis method for arbitrary ECG, UI-Beat, that can synchronously output the confidence of the diagnostic decision. By disentangling epistemic and aleatoric uncertainties, this method unifies noise screening, artifact or ambiguous heartbeat identification, heartbeat localization, and classification as a one-stage framework. It provides a feasible and deployable solution for heartbeat analysis for ECGs, especially for wearable ECGs with limited and nonstandard lead(s). Furthermore, the uncertainty-based cross-lead fusion significantly improves the performance in heartbeat localization and classification, indicating that UI-Beat can unify various forms of ECG monitoring into a core solution based on a general method for single-lead scenario. This presents a fruitful area for further work in migrating the UI-Beat to a broader application in long-term ECG diagnosis and other wearable cardiovascular signal monitoring scenarios.

References

[1] H. Chen and K. Maharatna, "An automatic r and t peak detection method based on the combination of hierarchical clustering and discrete wavelet transform," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2825–2832, 2020.

[2] M. Gabbouj, S. Kiranyaz, J. Malik, M. U. Zahid, T. Ince, M. E. Chowdhury, A. Khandakar, and A. Tahir, "Robust peak detection for holter ecgs by self-organized operational neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[3] N. Sabor, G. Gendy, H. Mohammed, G. Wang, and Y. Lian, "Robust arrhythmia classification based on qrs detection and a compact 1d-cnn for wearable ecg devices," *IEEE Journal of Biomedical and Health Informatics*, 2022.

[4] X. Wang, Y. Li, H. Gao, X. Cheng, J. Li, and C. Liu, "A causal intervention scheme for semantic segmentation of quasi-periodic cardiovascular signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3175–3186, 2023.

[5] C. Orphanidou and I. Drobnjak, "Quality assessment of ambulatory ecg using wavelet entropy of the hrv signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1216–1223, 2016.

[6] U. Satija, B. Ramkumar, and M. S. Manikandan, "A review of signal processing techniques for electrocardiogram signal quality assessment," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 36–52, 2018.

[7] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision," in *Neural Information Processing Systems*, 2017.

[8] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *International Conference on Machine Learning*. PMLR, 2016, pp. 1050–1059.

[9] R. Rahaman *et al.*, "Uncertainty quantification and deep ensembles," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[10] T. Zhu, A. E. Johnson, J. Behar, and G. D. Clifford, "Crowd-sourced annotation of ecg signals using contextual information," *Annals of biomedical engineering*, vol. 42, pp. 871–884, 2014.

[11] R. He, Y. Liu, K. Wang, N. Zhao, Y. Yuan, Q. Li, and H. Zhang, "Automatic detection of qrs complexes using dual channels based on u-net and bidirectional long short-term memory," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1052–1061, 2020.

[12] R. V. Andreao, B. Dorizzi, and J. Boudy, "Ecg signal analysis through hidden markov models," *IEEE Transactions on Biomedical engineering*, vol. 53, no. 8, pp. 1541–1549, 2006.

[13] K. Wang, G. Yang, Y. Huang, and Y. Yin, "Multi-scale differential feature for ecg biometrics with collective matrix factorization," *Pattern Recognition*, vol. 102, p. 107211, 2020.

[14] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, pp. 65–69, 2019.

[15] J. Malik, O. C. Devecioglu, S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1d self-operational neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 5, pp. 1788–1801, 2021.

[16] L. Sun, C. Li, Y. Ren, and Y. Zhang, "A multitask dynamic graph attention autoencoder for imbalanced multilabel time series classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[17] S. Zhang, Y. Li, X. Wang, H. Gao, J. Li, and C. Liu, "Label decoupling strategy for 12-lead ecg classification," *Knowledge-Based Systems*, vol. 263, p. 110298, 2023.

[18] N. Wang, P. Feng, Z. Ge, Y. Zhou, B. Zhou, and Z. Wang, "Adversarial spatiotemporal contrastive learning for electrocardiogram signals," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[19] W. Zhang, L. Yang, S. Geng, and S. Hong, "Self-supervised time series representation learning via cross reconstruction transformer," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2023.

[20] Z. Zhao, X. Wang, Z. Cai, J. Li, and C. Liu, "Pvc recognition for wearable ecgs using modified frequency slice wavelet transform and convolutional neural network," in *2019 Computing in Cardiology (CinC)*. IEEE, 2019, pp. 1–4.

[21] J. Niu, Y. Tang, Z. Sun, and W. Zhang, "Inter-patient ecg classification with symbolic representations and multi-perspective convolutional neural networks," *IEEE journal of biomedical and health informatics*, vol. 24, no. 5, pp. 1321–1332, 2019.

[22] H. Gao, X. Wang, Z. Chen, M. Wu, J. Li, and C. Liu, "Ecg-cl: A comprehensive electrocardiogram interpretation method based on continual learning," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[23] J. F. Vranken, R. R. van de Leur, D. K. Gupta, L. E. Juarez Orozco, R. J. Hassink, P. van der Harst, P. A. Doevendans, S. Gulshad, and R. van Es, "Uncertainty estimation for deep learning-based automated analysis of 12-lead electrocardiograms," *European Heart Journal-Digital Health*, vol. 2, no. 3, pp. 401–415, 2021.

[24] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 12, pp. 5586–5609, 2021.

[25] Y. Zhu, Y. Chen, Z. Lu, S. Pan, G.-R. Xue, Y. Yu, and Q. Yang, "Heterogeneous transfer learning for image classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 25, no. 1, 2011, pp. 1304–1309.

[26] Y. Li, X. Tian, T. Liu, and D. Tao, "On better exploring and exploiting task relationships in multitask learning: Joint model and feature learning," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 5, pp. 1975–1985, 2017.

[27] A. R. Zamir, A. Sax, N. Cheerla, R. Suri, Z. Cao, J. Malik, and L. J. Guibas, "Robust learning through cross-task consistency," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 197–11 206.

[28] F. Zhou, C. Shui, M. Abbasi, L.-É. Robitaille, B. Wang, and C. Gagné, "Task similarity estimation through adversarial multitask neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 466–480, 2020.

[29] W.-H. Li and H. Bilen, "Knowledge distillation for multi-task learning," in *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 2020, pp. 163–176.

[30] K. Li, J. Lu, H. Zuo, and G. Zhang, "Multi-source contribution learning for domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5293–5307, 2021.

[31] V. Aglietti, T. Damoulas, M. Álvarez, and J. González, "Multi-task causal learning with gaussian processes," *Advances in neural information processing systems*, vol. 33, pp. 6293–6304, 2020.

[32] J. Futoma, S. Hariharan, and K. Heller, "Learning to detect sepsis with a multitask gaussian process rnn classifier," in *International conference on machine learning*. PMLR, 2017, pp. 1174–1182.

[33] J. Liu, Z. Lin, S. Padhy, D. Tran, T. Bedrax Weiss, and B. Lakshminarayanan, "Simple and principled uncertainty estimation with deterministic deep learning via distance awareness," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7498–7512, 2020.

[34] L. Smith and Y. Gal, "Understanding measures of uncertainty for adversarial example detection," in *Uncertainty in Artificial Intelligence*. PMLR, 2018.

[35] J. Van Amersfoort, L. Smith, Y. W. Teh, and Y. Gal, "Uncertainty estimation using a single deep deterministic neural network," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9690–9700.

[36] J. Mukhoti, A. Kirsch, J. van Amersfoort, P. H. Torr, and Y. Gal, "Deterministic neural networks with appropriate inductive biases capture epistemic and aleatoric uncertainty," *arXiv e-prints*, pp. arXiv–2102, 2021.

[37] L. Kong, J. Sun, and C. Zhang, "Sde-net: Equipping deep neural networks with uncertainty estimates," in *International Conference on Machine Learning*. PMLR, 2020, pp. 5405–5415.

[38] J. Mukhoti, A. Kirsch, J. van Amersfoort, P. H. Torr, and Y. Gal, "Deep deterministic uncertainty: A new simple baseline," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 24 384–24 394.

[39] Z. Cai, C. Liu, H. Gao, X. Wang, L. Zhao, Q. Shen, E. Ng, and J. Li, "An open-access long-term wearable ecg database for premature ventricular contractions and supraventricular premature beat detection," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 11, pp. 2663–2667, 2020.

[40] H. Gao, C. Liu, X. Wang, L. Zhao, Q. Shen, E. Ng, and J. Li, "An open-access ecg database for algorithm evaluation of qrs detection and heart rate estimation," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 9, pp. 1853–1858, 2019.

[41] F. Liu, C. Liu, L. Zhao, X. Jiang, Z. Zhang, J. Li, S. Wei, and Y. Zhang, "Dynamic ecg signal quality evaluation based on the generalized bsqi index," *Ieee Access*, vol. 6, pp. 41 892–41 902, 2018.

[42] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter," *Physiological Measurement*, vol. 29, no. 1, p. 15, 2007.

[43] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[44] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg," in *Computers in Cardiology*. IEEE, 1997, pp. 673–676.

[45] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. Moody, C. Zeelenberg, and C. Marchesi, "The european st-t database: standard for evaluating systems for the analysis of st-t changes in ambulatory electrocardiography," *European Heart Journal*, vol. 13, no. 9, pp. 1164–1172, 1992.

[46] A. I. Kalyakulina, I. I. Yusipov, V. A. Moskalenko, A. V. Nikolskiy, K. A. Kosonogov, G. V. Osipov, N. Y. Zolotykh, and M. V. Ivanchenko, "Ludb: A new open-access validation tool for electrocardiogram delineation algorithms," *IEEE Access*, vol. 8, pp. 186 181–186 190, 2020.

[47] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 230–236, 1985.

[48] W. Cai and D. Hu, "Qrs complex detection using novel deep learning neural networks," *IEEE Access*, vol. 8, pp. 97 082–97 089, 2020.

[49] V. Kalidas and L. Tamil, "Real-time qrs detector using stationary wavelet transform for automated ecg analysis," in *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2017, pp. 457–461.

[50] M. U. Zahid, S. Kiranyaz, T. Ince, O. C. Devecioglu, M. E. Chowdhury, A. Khandakar, A. Tahir, and M. Gabbouj, "Robust r-peak detection in low-quality holter ecgs using 1d convolutional neural network," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 1, pp. 119–128, 2021.

[51] A. AAMI and A. EC57, "Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms," *American National Standards Institute, Arlington, VA, USA*, 2008.

[52] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE transactions on biomedical engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.

[53] J. He, L. Sun, J. Rong, H. Wang, and Y. Zhang, "A pyramid-like model for heartbeat classification from ecg recordings," *PloS one*, vol. 13, no. 11, p. e0206593, 2018.

[54] G. Wang, M. Chen, Z. Ding, J. Li, H. Yang, and P. Zhang, "Inter-patient ecg arrhythmia heartbeat classification based on unsupervised domain adaptation," *Neurocomputing*, vol. 454, pp. 339–349, 2021.

[55] J. Yang, W. Cai, and M. Wang, "Premature beats detection based on a novel convolutional neural network," *Physiological Measurement*, vol. 42, no. 7, p. 075003, 2021.

[56] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ecg classification," *Computers in biology and medicine*, vol. 99, pp. 53–62, 2018.