

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Electrocardiogram Recognition Based on Variational AutoEncoder

Shaojie Chen, Zhaopeng Meng and Qing Zhao

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.76434>

Abstract

Subtle distortions on electrocardiogram (ECG) can help doctors to diagnose some serious larvaceous heart sickness on their patients. However, it is difficult to find them manually because of disturbing factors such as baseline wander and high-frequency noise. In this chapter, we propose a method based on variational autoencoder to distinguish these distortions automatically and efficiently. We test our method on three ECG datasets from Physionet by adding some tiny artificial distortions. Comparing with other approaches adopting autoencoders [e.g., contractive autoencoder, denoising autoencoder (DAE)], the results of our experiment show that our method improves the performance of publicly available on ECG analysis on the distortions.

Keywords: electrocardiogram, variational autoencoder, variational inference, ECG enhancement, deep learning

1. Introduction

Automatic electrocardiogram (ECG) recognition [29] is greatly helpful to doctors in their diagnosis and treatment of heart disease. As the number of portable ECG devices is increasing, more and more ECG records are available. However, it is inevitable that these ECG data are contaminated by different kinds of noise caused by such interference as baseline wandering, muscle shaking, and electrode movement [13, 14]. Considering the level and complexity of these noises, especially those components that may cause subtle deformations on ECG waveforms, these factors may decrease the accuracy of the ECG recognition. Additionally, there are much more unlabeled ECG data (i.e., there are not any type information about the data) that

are stored in a lot of databases. Therefore, it is necessary to improve the performance of automatic ECG classification in unsupervised context by choosing proper models and algorithms.

In order to prevent noisy inference, many approaches of preprocessing or enhancement of ECG were successfully employed to remove the contaminations. Traditionally, most of these approaches are based on the filtering technology on frequency domain. Ziarani et al. and Konrad [15] eliminated the power line noise by extracting a specified component of a signal and tracking its variations over time. Alfaouri et.al. [16] and Dewangan et al. [17] employed wavelet transform method to isolate baseline wander and effectively detect and suppress the presence of power line interference in ECG. Although these filters can help suppress the high-frequency interference, they may drop out some useful information on the heart illness simultaneously. Because the frequency spectrum spreads not only low band but also high band. To overcome these drawbacks of filtering-based methods, some adaptive methods have been proposed. Abdelmounim et al. [18] applied adaptive algorithm to remove those noise that subsequently adapt to the wavelets selected by proper thresholding. However, the author also reported that this method had its own relative disadvantage that it had incapability of removing baseline wandering smoothly and effectively. Additionally, other technologies such as Fourier transform (FT) and empirical mode decomposition (EMD) were also employed for ECG preprocessing [19, 20]. FT maps the higher frequency components into the low area. Similarly, EMD separates different ECG components by proper intrinsic mode functions.

Feature extraction is another important procedure of ECG recognition. ECG features consists of amplitudes, intervals, and segments, which are shown in **Figure 1**. Each feature indicates certain activities of heart. For example, P wave represents atrial depolarization, it causes both atria to contract and pump blood to ventricles. Any distortion of P wave indicates malfunction of atrial appears.

Traditionally, the goal of ECG feature extraction is to extract all abovementioned features. As the amplitude of R wave is much larger than any others, many approaches based on the QRS complex detection have been proposed. Chan et al. [21] used a specific template to match the preferred ECG signals by the computation of the correlation between them. Krasteva and Jekova et al. [22] successfully implemented this method to evaluate the heart

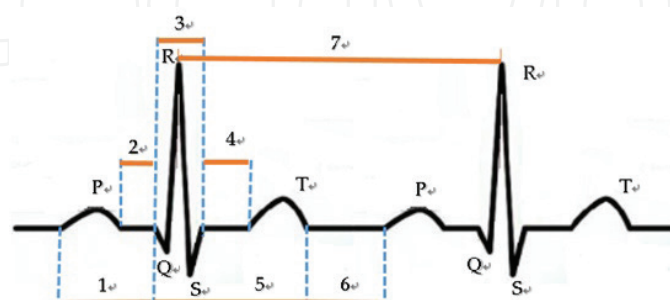


Figure 1. An ECG waveform with two cardiac periods. It consists of P wave, QRS complex, and T wave. Additionally, there are two intervals: PR interval (3) and QT interval (5). Three segments include PR segment (2), ST segment (4), and TP segment (6). RR interval (7) means how long is the duration between two adjacent peaks of R wave.

rhythm. Nevertheless, these approaches are heavily dependent on the prior knowledge about ECG and the relevant areas [23, 25], which cause more difficulties for further applications. Comparatively, some other approaches based on kernel functions are more popular and widely used because of their simplicity and sensitivity. Martis et al. [3] studied several methods [principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA), and discrete wavelet transform (DWT)] and compared them in feature extraction for classifying the arrhythmia ECGs. Banerjee et al. [5] focused on two specific regions (QRS complex area and T-wave region) on ECG waveforms to adequately distinguish between normal and abnormal ECG signals by yielding wavelet cross spectrum and wavelet coherence. Kærgaard et al. [6] proposed two hybrid signal processing schemes [ensemble empirical mode decomposition (EEMD) and discrete wavelet transform (DWT)] for ECG features extraction. These schemes were implemented by combining with the neural network and the wavelet transform. Nazarahari et al. [8] chose wavelet functions (WFs) as means of ECG classifying and proposed a wavelet design criterion for wavelet function choosing. Houssein et al. [4] classified the ECG by modified water wave optimization (WWO) algorithms and achieved over 93% average accuracy.

Although many important contributions have been given to ECG feature extraction by conventional methods based on kernel technologies, the accuracy and efficiency of these methods could rarely meet all the requirements of applications especially in the background of noise. Fortunately, different from the kernel methods, neural networks have been used to draw ECG features automatically by the hierarchical structure in the context of deep learning, which could be achieved by a new approach which is known as representation learning. Yan et al. [12] used a restricted Boltzmann machine (RBM) for ECG classification. Xiong et al. [9, 10] employed denoising autoencoder (DAE) and stacked contractive denoising autoencoder for ECG denoising [8], respectively. Zhou et al. [11] chose a stacked sparse autoencoder (SAE) to extract ECG feature for classifying and the level of accuracy achieved by this work shows derivable benefits over the traditional methods that require wavelets transform to perform ECG classification.

In terms of the heart illness automatically diagnosis auxiliary by the ECG recognition, some works mentioned above do not meet the necessary requirements because most studies focused on the arrhythmia distinguishing problems. Nevertheless, many heart diseases have close relationship not only with the rhythms of itself but also with the other features such as the length of the ST segment and the amplitude of P wave on the ECG waveforms. Additionally, there are rarely generative models to be used for ECG recognition. The contributions of this chapter include two aspects: (1) instead of using ECG signals on a cardiac period between two start points at P waves, we propose a new method for intercepting ECG segments between adjacent two R peaks and (2) we use variational autoencoder (VAE) model as an analysis tool to recognize different ECG signals by focusing on the variation of tiny distortion.

This chapter is organized as follows. Section 2 briefly describes autoencoder and its variants. Section 3 introduces the variational inference and variational autoencoder in detail. ECG preprocessing and classifying schema is proposed in Section 4. Our experiment results and discussions are shown in Section 5. Finally, Section 6 concludes.

2. Autoencoders and variants

Variational autoencoder has close relationship with autoencoder. An autoencoder is a neural network that consists of encoder and decoder. Encoder maps its input into representation and decoder reconstructs the representation back into the input, that is, perfect autoencoder can resemble the training data approximately by forcing to prioritize those aspects of the input that are helpful to resembling and discard the others. In this regard, the autoencoder learns the useful properties of training data. Comparatively, VAE shares the same character with AE besides some specialties of its own.

2.1. Autoencoder and regularized variants

Autoencoder can be used to get useful features from the encoder output. Generally, in the view of the feature dimension, autoencoder falls into two categories: undercomplete and overcomplete. Undercomplete means the dimension of feature is less than that of the input and more salient features could be learned well in this scenario. Conversely, in the case of overcomplete, the dimension of feature is greater than that of the input and more sparsity features might be drawn in this setting. Additionally, the objective function is another core topic for an autoencoder. It is designed to make the autoencoder have capabilities such as linear regression or logistic regression, which limit the model to some useful properties of the training data. The general form of the objective function can be depicted as follows:

$$\tilde{J}(X, \theta) = J(X; \theta) + \alpha \Omega(\theta). \quad (1)$$

where X is the training data for a given autoencoder. $\theta = \{W_e, b_e, W_d, b_d\}$ are the parameters of the model and α is a nonnegative hyperparameter that controls how much of the penalty term Ω to the relative to the standard objective function J . Numerically, setting α to 0 means not any regularization and larger values of α result in more regularization. Conceptually, autoencoders with penalty term is usually called regularized autoencoder that is encouraged to have small derivative of the representation, which leads the convergence faster than those that have not any regularization during the training time.

Varied forms of regularizer terms make the autoencoder have different properties and bring us different variants of regularized autoencoder. These variants include primarily sparse autoencoder (SAE), denoising autoencoder (DAE) [3], contractive autoencoder (CAE), and variational autoencoder (VAE). Theoretically, VAE combines variance inference (VI) and neural networks. As a generative model, one of the prominent successes of VAE is that it realizes effective random sampling using back-propagation (BP) technology. This will be described in detail in Section 3.

Different from VAE, SAE makes majority of the neurons in its hidden layers be inactive since the active functions on these neurons are feasibly saturated for most input. This results in the sparsity of features, where many of the elements of the features are zero (or close to zero). In the view of mathematics, the sparsity of SAE is accomplished by the penalty term $KL(\tilde{p}||p)$,

where \tilde{p} is the given sparsity value. Parameter p will be adjusted gradually to \tilde{p} in the training stage and achieve satisfactory sparsity. Analogous to SAE, CAE [4, 26] yields the specialized contractive properties by the penalized term—a Jacobin matrix that is consisted of the partial derivatives of the decoder active functions to input vectors. Then the input perturbations can be resisted during training time. Consequently, neighborhood of points in samples is encouraged to map into a smaller area, which can be thought as the capability of contracting for CAE. The motivation of DAE is to be insensitive to noise. Instead of adding an additional penalty term to the object function, DAE is trained by the noise-corrupted data \tilde{x} ($\tilde{x} = x + \beta\tau$) [27, 30]. DAE yields great success in many cases especially in manifold assumption. As the corrupted data \tilde{x} lie farther away from the manifold than the uncorrupted ones, DAE tends to take those points that are farther from the manifold to near. The larger distance from the manifold, the bigger step DAE takes to the manifold.

Generally, these autoencoders share some properties. DAE and CAE are able to learn the manifold structure of the samples. Simultaneously, SAE and CAE have the similar sparsity character on their representation. Nevertheless, the implementations of these autoencoders are quite different. For example, DAE reaches the goal by using the noise-corrupted data to train the structure to learn the proper parameters that can reconstruct the original samples without any noise. Comparatively, CAE takes Jacobian matrix as part of the loss function and encourages robustness on the representation by contracting the samples during the training process.

3. Variational inference and variational autoencoder

As the central problem in inference analysis, posterior distribution computation is facing two computing challenges: marginal likelihood computation and predictive distribution computation. Both of them are intractable since they often require computing high-dimensional integrals. Therefore, approximate inference approaches such as Gibbs sampling based on Markov chain Monte Carlo (MCMC) principle are appealing. However, Gibbs sampling and its variants are often restricted from some applications for their inefficiencies especially in the high-dimensional scenario. This awkward situation has not been changed until the VAE was proposed theoretically [36]. To get an understanding of a VAE, we will first start from the relevant bases including variational inference (VI), evidence low boundary (ELBO), mean field, and Kullback–Leibler (KL) divergence.

To describe the problem mathematically, let $x = \{x_1, x_2, \dots, x_N\}$ be a set of N observations and $Z = \{z_1, z_2, \dots, z_m\}$ be the m latent variables. $P(Z, X; \theta)$ denotes the joint distribution of x and z given the parameter θ of the model. $P(X|Z)$ and $P(Z|X)$ are called the likelihood of z and the posterior distribution of x , respectively.

3.1. Variational inference

Theoretically, the motivation of variational inference [33, 35] is to find a feasible distribution to approximate the desired posterior distribution that is intractable. To measure how

closeness of these two distributions are, Kullback–Leibler (KL) divergence [34] is introduced. Let $P(X)$ and $Q(X)$ indicate two different distributions of the continuous random variables X , their KL divergence is defined as:

$$\text{KL}(P||Q) = \int Q(X) \log \frac{Q(X)}{P(X)} dX = E_{Q(X)} \left[\log \frac{Q(X)}{P(X)} \right]. \quad (2)$$

Intuitively, KL divergence is nonnegative and monotonically decreasing to the similarity of the distributions, that is, the more similar of the two distributions, the smaller the KL divergence value is. The identity equals zero when $Q(X)$ is the same as $P(X)$. However, the KL divergence is non-symmetrical as $\text{KL}(Q||P) \neq \text{KL}(P||Q)$. The definition indicates implicitly another two properties: the KL divergence equals zero when $Q(X)$ goes infinitively to zero regardless of $P(X)$ and rises asymptotically infinity as $P(X)$ becomes zero. Hence, we can approximate the distribution $P(X)$ for $Q(X)$ by minimizing $\text{KL}(Q(X)||P(X))$.

3.1.1. Evidence lower boundary

In the context of Bayesian statistics, “Evidence” is an alternative term used for the marginal likelihood of the observations. Formula (3) reveals the relationship between KL divergence and the logarithm of the evidence $P(X)$. The difference between them equals the expectation of $\log(p(X, Z)) - \log(q(Z))$, which is called the evidence lower boundary (ELBO). As the KL divergence is nonnegative, then we have the evidence lower boundary as formula (3). Jordan et al. [1] got the same result originally using the Jensen’s inequality. Formula (3) shows literally the name of ELBO. We may define the expectation of $\log(p(X, Z)) - \log(q(Z))$ as $L(Q)$, a function of distribution of $Q(Z)$:

$$\log(P(X)) - \text{KL}(Q(Z)||P(Z|X)) = E_Q[\log(P(Z, X)) - \log(Q(Z))] \triangleq L(Q). \quad (3)$$

Intuitively, maximizing ELBO is equivalent to minimizing the KL divergence. As the $\text{KL}(Q(Z)||P(Z|X))$ decreases to zero, it is necessary to make the posterior distribution $P(Z|X)$ share the same distribution with $Q(Z)$. Hence, we can use $Q(Z)$ to approximate the posterior distribution $P(Z|X)$ by maximizing ELBO, which can be realized by optimizing the objective of $L(q)$ as formula (4), finding an optimal distribution $Q^*(Z)$ within a specifying family \mathcal{Q} of densities over the latent variables. Expectation maximization (EM) algorithm [2] is one of the successful approaches that were designed for finding the optimal solution $Q^*(Z)$ within the family \mathcal{Q} . It alternates iteratively between expectation step (E-step) where the posterior distribution $P(Z|X; \theta)$ is calculated and then, maximization step (M-step) where the expectation of the complete-data likelihood with respect to the posterior distribution $P(Z|X; \theta^{\text{old}})$ is maximized by optimizing the parameters θ^{new} . Then updates the parameters θ^{old} with θ^{new} :

$$Q^* = \underset{Q \in \mathcal{Q}}{\operatorname{argmax}} L(Q). \quad (4)$$

3.1.2. Mean field

To simplify the optimization problem of ELBO, it is necessary to make assumption on the family \mathcal{Q} , as the selection of the family affects impressively on complexity of the optimization algorithm for the problem. This assumption focuses on the way that how to factorize $Q(Z)$ as:

$$Q(Z) = \prod_{i=1}^m Q_i(Z_i). \quad (5)$$

where $Q_i(Z_i)$ denotes the individual factors that are mutually independent over the latent variables of the model. According to the chain rule of probability, the joint distribution $P(X, Z)$ can be decomposed as:

$$p(Z, X) = p(X) \prod_{i=1}^m p(Z_i | Z_{1:(i-1)}, X). \quad (6)$$

Then, the ELBO can be written as Eq. (7):

$$L(Q) = \log(P(X)) + \sum_{i=1}^m \left(E_Q \left[\log \left(P(Z_i | Z_{1:(i-1)}, X) \right) \right] - E_{Q_i} \left[\log(Q_i(Z_i)) \right] \right). \quad (7)$$

where $\log(P(X))$ is constant with respect to $Q(Z)$. Then, maximizing ELBO is equivalently maximizing the last summation term. Furthermore, we can derive out the optimal solution Q^* by Lagrangian multiplier method:

$$Q_i^*(Z_i) \propto \exp \left(E_{-Q_i} \left[\log(Z_i | Z_{-i}, X) \right] \right). \quad (8)$$

Formula (8) indicates that the factors are all proportional to the exponentiated log the joint distribution except the i^{th} variational factor. This is the gist of the coordinate ascent variational inference (CAIV) [37] as well. However, as the ELBO is not a necessary convex function, there is no guarantee that the solution Q^* is a global optimum.

3.2. Variational autoencoder

As a deterministic model, general regularized autoencoder does not know anything about how to create a latent vector until a sample is input. Conversely, as a generative model, variational autoencoder (VAE) [36] emerges as a successful example of combination of variance inference and neural network. VAE forces the latent vector following some kind of distribution. These characters not only encourage the properties of the general regularized autoencoders but also expand some additional properties. For example, VAE can generate some data points even without any encoding input. It is the specialty of VAE that differs from the other regularized autoencoders. To explore VAE further, it is necessary to understand those complicated ideas such as the neural network structure, the loss function, and the optimization algorithm.

In the view of the hierarchy, the neural network structure of the VAE is mainly composed of three parts. The first part is the encoder, which is used to encode the signals from the input layer. The second part is the decoder, which is located in the right side as shown in **Figure 2**. The third part is the sampling unit located in the middle of the other two parts. Except for the encoder and the decoder which are similar to that of the traditional autoencoder, the additional sampling unit is responsible for sampling from the latent variables spaces.

Another issue about how to train the structure is the loss function as shown in formula (9), which is essentially the same as the negative $L(Q)$ in formula (7). In the view of training, the losses of a VAE come from two aspects: the first part is from the neural network that measures how much the difference between the reconstructed data and the original input. This part encourages the decoder to learn to reconstruct the input. Otherwise, the value of this part will become even larger that will increase the total loss value finally. The second part comes from the KL divergence that indicates how much close of the encoder's distribution $Q(Z|X)$ and the latent variables distribution. This part can be taken as a regularizer as that of the traditional autoencoder. It forces the encoder's distribution $Q(Z|X_i)$ go as close to the latent variables distribution $P(Z)$ as possible by minimizing KL divergence of them. In other words, if the encoder outputs representations are different from the specified distribution, then the regularizer term will penalize the loss function. Otherwise, the penalty will vanish away:

$$L_{VAE} = -\sum_{i=1}^N \left(E_{Z \sim Q(Z|X_i)} [\log(P(X_i|Z))] - KL(Q(Z|X_i) \| P(Z)) \right). \quad (9)$$

The last idea for VAE is the way that how to minimize the loss function of Eq. (9) as working on the neural networks, where the algorithms based on gradient decent are popularly adopted. Comparatively, it is feasible to compute the first term in the Eq. (9) as the expectation indicates the reconstruction difference and we can calculate it by the mean squared error between the output of the encoder and the decoder, as similar to that of the traditional autoencoders. However, it is more difficult to compute the second KL divergence directly as $P(Z)$ and $P(X_i|Z)$ are all intractable. Fortunately, An effective solution was proposed by Kingma et al. [36] on the assumption that $Q(Z|X_i)$ follows a normal distribution $Q(Z|X^i) \sim \mathcal{N}(Z; \theta)$, where $\theta = \{\mu_1, \Sigma_1\}$ and μ_1 and Σ_1 are the parameters of the mean and the variance, respectively. For the simplicity, here we assume $P(Z) = \mathcal{N}(Z; 0, I)$, where I is a unitary diagonal matrix. The advantages of this choice make the computation of the KL divergence manageable. We can compute it in the closed form as:

$$KL(Q(Z|X_i) \| P(Z)) = \frac{1}{2} \left(\log \frac{1}{|\Sigma_1|} - D + \text{tr}(\Sigma_1) + (\mu_1)^T (\mu_1) \right). \quad (10)$$

D is a constant value that is only relevant to the dimensionality of the distribution.

Additionally, to train a VAE neural structure, the gradient decent should be focused on when error back propagates through the sampling layers. However, we cannot derivate the loss function over the distribution $Q(Z|X_i)$ directly as the distribution is a non-continuous operation and has no gradient. To clarify the problem, suppose we can take the derivation of J_{VAE} respect to $Q(Z|X_i)$, then we get the gradient expression as following:

$$\frac{\partial L_{VAE}}{\partial Q} = \log(P(X_i|Z)) + \log(Q) - \log(P(Z)) + \text{const}. \quad (11)$$

It is clear that the gradient depends not only on the decoder's distribution $P(x_i | z)$ but also on the encoder's distribution $Q(z | x_i)$. Except for the non-continuity of the encoder's distribution, there is no stochastic unit with the neural network. Kingma et al. [36] presented a method named "reparameterization trick" to solve the problem successfully. Instead of drawing from the encoder's representations directly, sampling unit generates μ and σ at first by sampling from the input x . Given $\mu(x)$ and $\sigma(x)$, we can do sampling from $\mathcal{N}(\mu(x), \sigma^2(x))$, and then compute $z = \mu(x) + \sigma(x) * \epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$. Consequently, given a fixed x and ϵ , L_{VAE} becomes continuous and deterministic for P and Q , which means that derivation of L_{VAE} over Q is computable. Then those algorithms based on the gradient descent (GD) can be effective on VAE neural networks. Comparing to the time-consuming Gibbs sampling methods, algorithms based on GD are much more effective and efficient.

4. ECG preprocessing and enhancement

In this section, we introduce our method on ECG preprocessing and enhancement. The task in this procedure is to split the ECG waves into segments according to the cardiac cycle [28] and then take them as data points for training our models. As described in Section 1, QRS complex is responsible for the activities of ventricular depolarization and repolarization, it has morphologically higher amplitude and sharper peak than other components such as P-wave and T-wave. Therefore, it is much more convenient to detect and locate Q peaks (or R, S peaks) than any other components in these ECG segments. Algorithm 1 describes the procedure of how to split ECG waveforms in detail. The templates selected in algorithm 1 are produced by the contours of the most ECG R wave peaks.

The critical step in Algorithm 1 is how to evaluate the similarity between the selected area on the ECG waveform and the given template. Generally, the mean squared error (MSE) is usually adopted in some ECG recognizing applications. However, the main disadvantage of this method is that it is time-consuming to align the selected area with the given template. For example, there are two pictures with the same curve, the similar value of the pictures may be definitely tiny if the template aligns extremely well or a very large as they do not cover each other at all. Another reasonable approach named the correlation coefficient is being currently used [21, 26]. Instead of computing directly the difference between the ECG waveform and the template as the MSE method, it solves an optimal problem that minimizes the sum of the squares of the offsets of the selected ECG data points to the corresponding points on the template.

We introduce a parameter h_{step} for the length of the segment of ECG waveforms. It is important to keep h_{step} lie in a proper range. Otherwise, there are more than one R peaks or none in the segment when the h_{step} is out of the range. To avoid the awkward situations, there is a trick that let the h_{step} be proportional to the distance between two adjacent peaks and rather less than it, that is, $h_{step} \leq \frac{\text{sampling rate}}{\text{heart rate}}$. For instance, suppose sampling rate is 250 Hz and heart rate equals 75 times per minute, then $h_{step} \leq 200$. As the heart rate is not a constant during the sampling procedure, then distance can be calculated by the inequation. For this reason, in all of our experiments, the distance is set empirically as the average of that of previous three cardiac periods. The searching step can be initialized as a constant value as there are no any variations on the vertical directions. We keep the v_{step} equaling 1 in this chapter.

Algorithm 1. ECG R wave peak location algorithm.

```

1: input: ECG data file name pa

2: initial: set segment length hstep and searching step vstep, empty ECG data buffer ecg_v[M] and R wave peaks array ecg_pos[m];

3: read ECG data into ECG data buffer ecg_v from ECG data file ecg_data_file;

4: calculate segment number  $N = \lceil L/hstep \rceil$  where  $L = \text{length}(ecg_v)$ ;

5: for each segment s in N

6:   let search range in vertical direction equal start position;

7:   while not bfind and tp > 0 and bp > 0 do

8:     Look for R wave peak in small area of [rp lp] in range of [tp, bp] using template;

9:     if isfindpeak() // to decide whether find the target.

10:      Save the result to ecg_pos;

11:      break;

12:   else

13:     Update range of [tp, bp] for next iteration;

14:   end if

15: end while

16: update rp and lp respectively;

17: end for

18: return ECG data array v, R wave peak array ecg_pos;

```

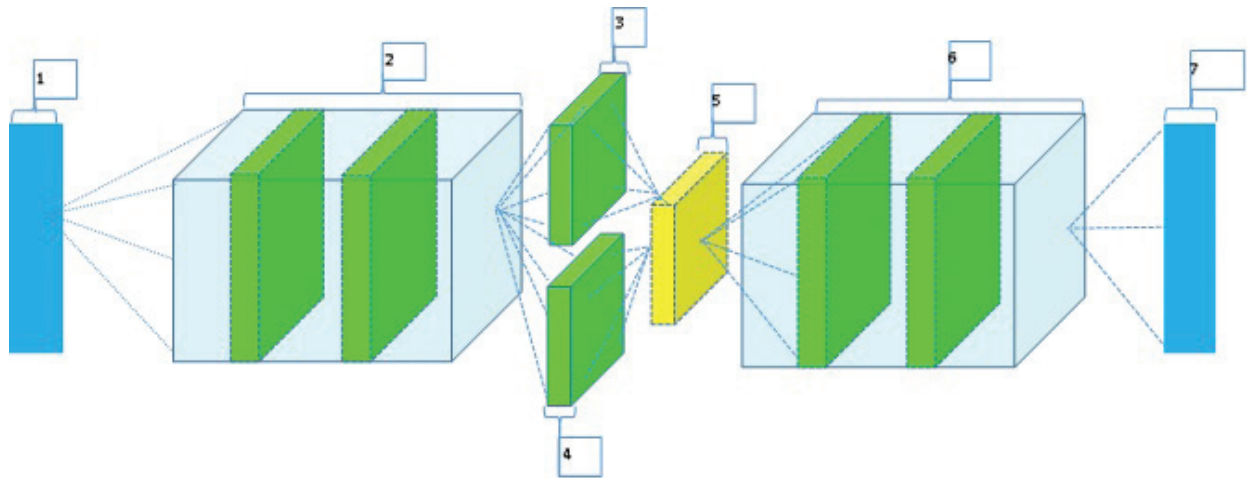


Figure 2. Neural network structure of VAE. It consists of three parts: The encoder, the decoder, and the sampling unit. The encoder (indicating by number 2) and the decoder (indicating by number 6) are all fully connected multilayers neural networks. The sampling unit consists of the mean generator (indicating by number 3), the standard deviation generator (indicating by number 4), and the latent vector generator (indicating by number 5). The structure of the sampling unit lies on the assumption of $Z \sim N(\mu, \sigma^2)$.

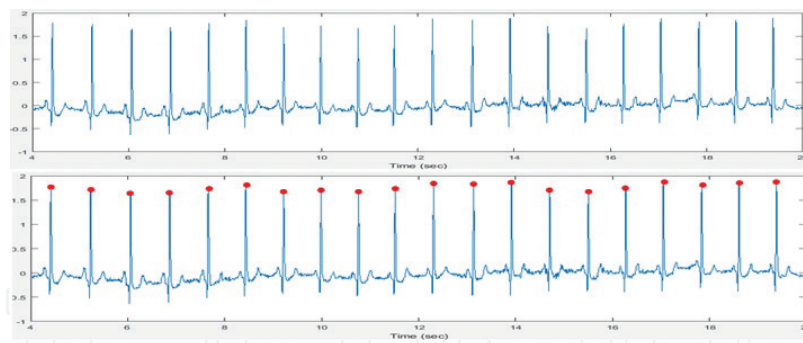


Figure 3. ECG waveform and R-wave peak location adopted from AHA database (top). The bottom picture shows the result of R peaks detection and location for the ECG waveform in the top picture.

Figure 3 shows ECG waveform (top picture) and the R wave peak detection and location (bottom picture). The ECG data are adopted from the American Heart Association (AHA) database on physionet website [24], which consisted of 80 two-channel ECG recordings and digitized at 250 Hz with 12-bit resolution over a 10-mV range. The recordings in the database are divided into eight classes according to the highest level of ventricular ectopy present.

5. Experimental results and discussion

In this section, we evaluate the performance of VAE and other autoencoder variants described in Section 2.

5.1. ECG signals for multi-classification

To demonstrate the performance of our models on dealing with ECG signals, it is necessary to abstract an intact ECG signal in a cardiac period, which consists of features such as P-wave, QRS complex, and T-wave as described in Section 4. Then detection and location of P-wave becomes more critical step as every cardiac period of ECG signal starts at P-wave. However, as the amplitude of P-wave is smaller than that of QRS complex, and there are many kinds of noise on ECG singles. These factors enlarge the difficulties of abstraction of ECG signals in a cardiac period.

Our solution to alleviate this problem is offered by the fact that it is more feasible to locate R-peaks than to locate the start position of a P-wave. Instead of focusing on the cardiac period, we separate one cardiac period into two semi-cardiac periods at R-peak and then take two parts of the adjacent ECG signals together to form a new period ECG signal, which consists of the second part of the previous cardiac period and the first part of the next one. **Figure 4(a)** shows an example of an ECG signal that is composed of two parts of the adjacent semi-period. Additionally, in the view of information, there is no any feature lost in this separation.

The original ECG recording from ECG database contains several hours of ECG data, and it is unfeasible to train our models using these original ECG data directly. To train our models well, 30,000 ECG signals are abstracted completely from three different ECG databases.

The AHA ECG database, the APNEA ECG database [24], and CHFDB ECG database [24]. Additionally, for ECG data augmentation [32], these ECG data are divided into three different groups according to their source databases and each group has 10,000 ECG signals. On this basis, we augment the ECG data by zeroing a small segment on ECG signals and different positions we selected to zero correspond to different class labels. **Figure 4(b)–(d)** are three examples of our augmentation. Concretely, the labels of **Figure 4(b)–(d)** are 3, 4, and 5, respectively. (We use numbers 1–8 as eight labels for different class of ECG signals in all of our experiments. We add labels for the different classes of ECG signals, not for training our models but for simplifying evaluating the accuracy of our models in testing process.)

To evaluate the properties of our models on denoising for ECG signals, different type noise on different level are added into the original ECG records. These noise include Gaussian noise, salt and pepper noise, and Poisson noise. Moreover, to imitate baseline wandering noise, different amplitude sinusoidal signals are superimposed on the original ECG signals. The coefficients of the sinusoidal signal are 0.01, 0.05, and 0.1, respectively in all of our experiments. **Figure 5** shows the ECG signals polluted by different noises. **Figure 5(a)** and **(c)** show the augmented ECG signals without adding noise except for some one polluted during sampling. **Figure 5(b)** shows ECG signal polluted by the sinusoidal noise and the Gaussian noise. The coefficients for the sinusoidal and for the Gaussian are all 0.01. Nevertheless, the coefficients for the sinusoidal and for the Gaussian are 0.05 and 1 as shown in **Figure 5(d)**. The mean and variance of the Gaussian noise are 0 and 0.01, respectively.

5.2. Recognition of ECG signals

After ECG signals have been abstracted completely by the methods described in Section 5.1, they are used to train VAE model. To compare the effect of the complexity of ECG data on our model, all ECG data are divided into two groups. The first one contains only two classes of ECG records, normal or abnormal. (We call this group as BI dataset) The normal ECG records mean those ones that contain all normal features as shown in **Figures 4** and **5**. The abnormal ECG records in BI dataset contain at least one abnormal feature such as prolonged PR interval, enlarged P-wave, and absence of T-wave. The second group contains 8 classes of ECG records, each of them are produced by zeroing a small segment of ECG

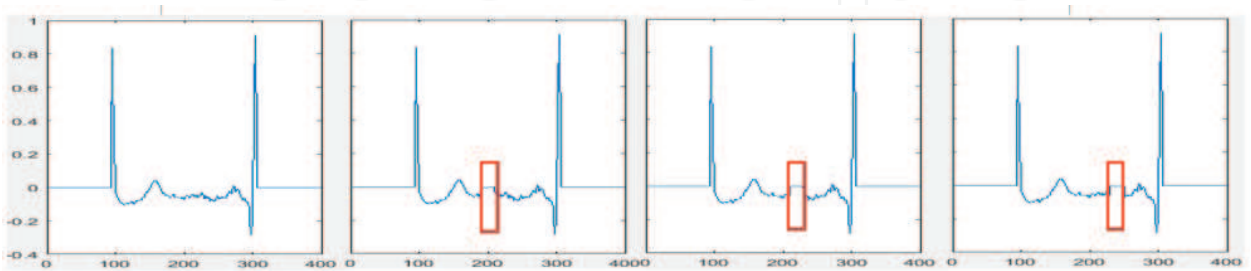


Figure 4. An example of ECG signals that is composed by two parts of the adjacent semi-period. (a) Single period ECG signal between the adjacent R peaks derived by algorithm 1. (b)–(d) are different class ECG signals derived by making a small segment of the same ECG signal zero on different position. The difference is marked by red rectangle area.

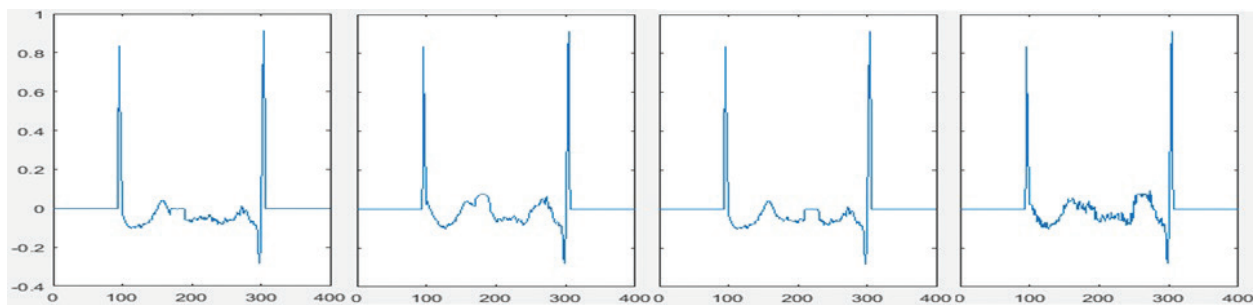


Figure 5. Single periodic ECG signal polluted by different noises. (a) Original ECG signal without adding noise. (b) ECG signal of (a) with Gaussian noise. (c) Original ECG signal with a segment- flatness. (d) ECG signal of (c) contaminated by Gaussian and sine wave noise imitating basing line wander.

data as described in Section 5.1 (We call this group as MI dataset). In order to verify the performance of the VAE model on ECG signals, the parameters of the model are shown in the **Table 1**. **Table 2** shows the performance of the VAE model on recognizing these ECG signals from both BI and MI datasets. The results clearly show that the accuracies of recognition are higher than 95% for MI recorders and even more than 97% for BI recorders. In the view of the data complexity, the result is reasonable because the complexity of MI is much higher than that of BI.

Advantages of VAE model on recognition ECG signals can be further shown by comparasion with other autoecoders such as CAE,DAE, and SAE mentioned in Section 2. In order to make the comparison be fair and reasonable, all of the parameters of the model are the same exept for that of the sampler in VAE model (the values of the parameters can be seen in **Table 1**). Moreover, the ECG records of BI and MI from ahadb database are used to train and test all the models. **Figure 6** shows the accuracy of the models on recognizing the ECG records. Both **(a)** and **(b)** in **Figure 6** take the rate of the representation to the input on size as variable. **Figure 6(a)** takes the BI ECG records from the ahadb as the datasource for the models. Conversely, the MI

Parameter name	Value	Comment
Input size	400	Equal the length of signal
h1	100	First layer of the encoder
h2	10	Second layer of the encoder
z-mean	2	Mean of the sampler
z-variance	2	Variance of sampler
Learning rate	0.01	
Function	Log-sigma	Logarithmic sigma
Optimizer	AdamOptimizer	
Batch size	100	Randomly select samples from the dataset

Table 1. Parameters of VAE model.

DB	Record	ECG no.	Sample no. (10 ³)	Class no.	Precision (%)	Error (%)
ahadb	0001	0	10	2	97.70	2.30
ahadb	0001	0	10	8	96.31	3.69
ahadb	0001	1	10	2	96.63	3.37
ahadb	0001	1	10	8	93.91	6.09
ahadb	0201	0	10	2	99.87	0.13
ahadb	0201	0	10	8	96.58	3.42
ahadb	0201	1	10	2	98.10	1.90
ahadb	0201	1	10	8	98.25	1.75
APNEA	a01	0	0.7	2	98.02	1.98
APNEA	a01	0	0.7	8	97.56	2.44
APNEA	a02	0	0.8	2	99.87	0.13
APNEA	a02	0	0.8	8	95.74	4.26
CHFDB	Chf01	0	10	2	99.99	0.01
CHFDB	Chf01	0	10	8	97.65	2.35
CHFDB	Chf01	1	10	2	98.89	1.11
CHFDB	Chf01	1	10	8	96.45	3.55
CHFDB	Chf01	0	10	2	99.75	0.25
CHFDB	Chf01	0	10	8	96.78	3.22
CHFDB	Chf01	1	10	2	99.26	0.74
CHFDB	Chf01	1	10	8	97.92	2.08

Table 2. Performance evaluation of VAE model on three ECG databases.

records from the same dataset are selected in **Figure 6(b)**. It is clear that the accuracy of the VAE model is higher than that of the other models on both BI and MI ECG records, which is at least 95% on BI records and no more than 90% on MI records. Meanwhile, both figures indicate a fact that the proper rate for the accuracy on the same condition is at 1. The accuracy is near 80% when rate falls at 0.5. Similarly, the accuracy drops sharply as the rate rises up. Therefore, there is no necessary for representation of ECG signals to compress (rate < 1) or stretch (rate > 1) themselves.

Figure 7 demonstrates the performance of the VAE model on denoising for ECG records. The method of adding noise into ECG records in our experiment can be seen in Section 5.1. The coefficient for sinusoidal is 0.05 and the mean and the variance of Gaussian noise are 0 and 0.05, respectively. For the goal of comparison, we take four groups of ECG records (BI, noisy BI, MI and noisy MI) as dataset for the VAE model.

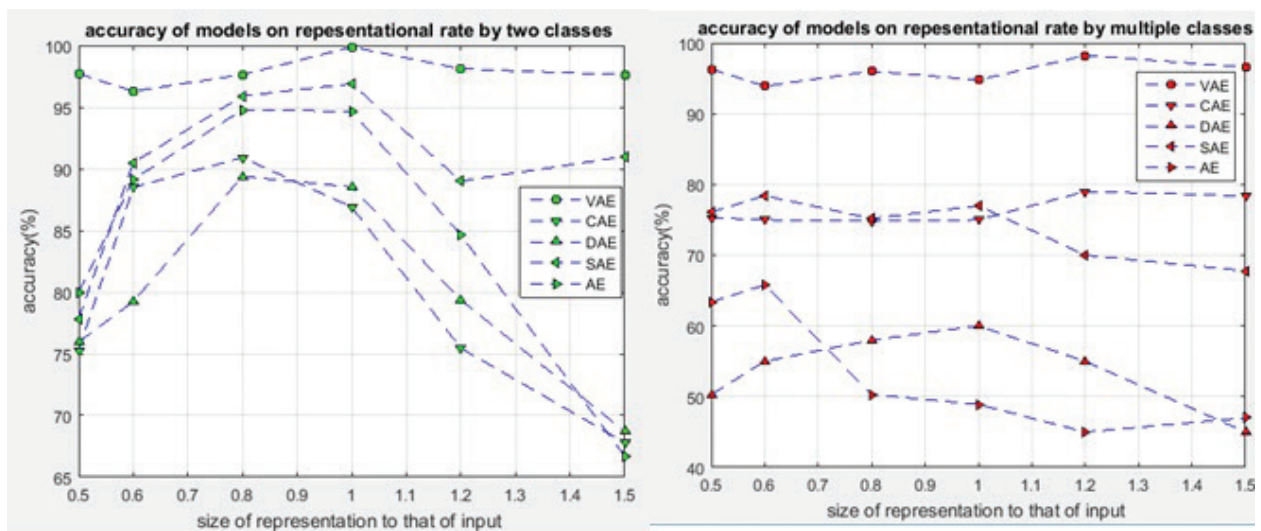


Figure 6. Accuracy of different models on recognition ECG signals from aha database. (a) Accuracy of the models on recognizing ECG signals from BI dataset of ahaadb ECG database. (b) Accuracy of the models on recognizing ECG signals from MI dataset of ahaadb ECG database.

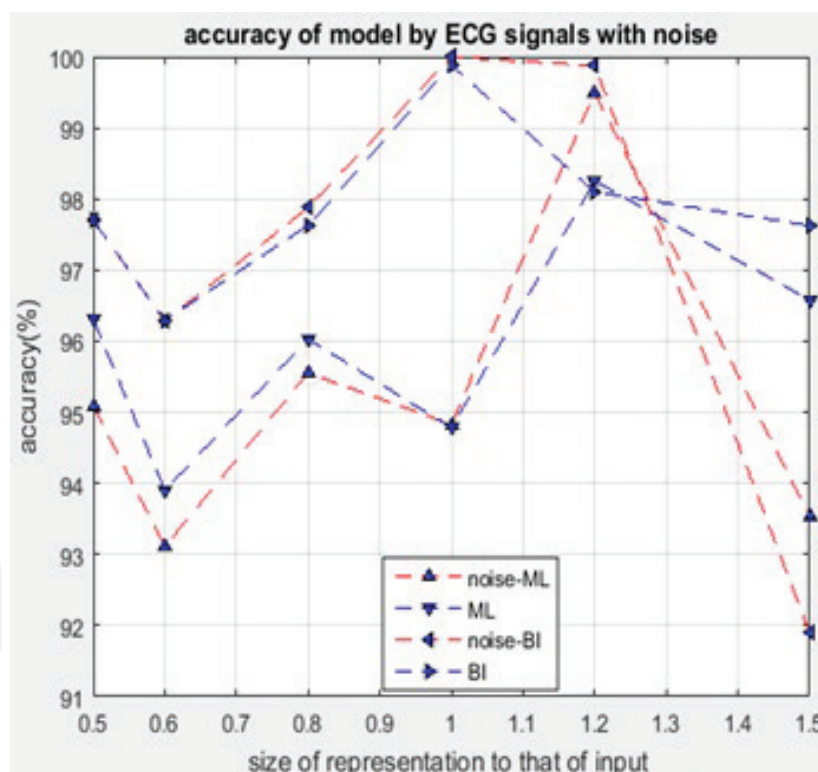


Figure 7. The performance of the VAE model on denoising for ECG records.

The results show that the accuracy under noisy condition is similar to that of without noise on the same dataset. This means that performance of VAE model on ECG recognition is robust to some kinds of noises.

6. Conclusions

In this chapter, we develop a VAE model to recognize a tiny distortion on ECG signals. First, we analyze the characteristics of the features of the ECG signals, which are closely related to ECG components such as P-waves, QRS complex, and T-waves. Second, we explain an algorithm that deals with the location of R peaks. On the basis of the algorithm, we abstract a segment of ECG signal between two adjacent R peaks from three real-life ECG databases. Finally, we train our models by using the selected ECG signals. The results of our experiments demonstrate that the proposed VAE model can be used as an effective tool to automatically recognize ECG signals. Especially, this model is robust to some kinds of noises that are usually produced during the sampling procedures. Furthermore, as a generative model, VAE is a recently established based on the neural networks. The important characteristic of the model is that it can be used in the scenario of the unsupervised learning [31]. Simultaneously, with the emergence of the large amount of unlabeled ECG records and the requirement for real-time diagnosis of heart illness by automatic recognition ECG signals, our method in this chapter can offer a solution to these problems.

In the view of the clinic, future work should put more energy on setting up the set of features of ECG signals, especially, the relationship between the features and the heart diseases. Additionally, because of the physiological characteristics of heart, a single ECG wave may not accurately represent the entire situation of the heart, it is therefore desirable to obtain all of ECG signals from all of 12 or 18 leads. For example, if an anterior wall myocardial infarction happens. Feature of ST-segment elevation reciprocally changes on the ECGs from the leads of I, aVL, and V1–V5. Therefore, the general implementation of VAE model to such clinic situations warrants further study.

Author details

Shaojie Chen^{1,2}, Zhaopeng Meng¹ and Qing Zhao^{2*}

*Address all correspondence to: zhaoqing@tust.edu.cn

1 School of Computer Software, Tianjin University, Tianjin, China

2 School of Computer Science and Information Engineering, Tianjin University of Science and Technology, Tianjin, China

References

- [1] Jordan MI, Ghahramani Z, Jaakkola TS, et al. Introduction to variational methods for graphical models. *Machine Learning*. 1999;37(2):183-233
- [2] Bishop, Christopher M. Pattern recognition and machine learning. In: *Information Science and Statistics*. New York: Springer-Verlag; 2007. p. 049901

- [3] Martis RJ, Acharya UR, Min LC. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomedical Signal Processing and Control*. 2013;**8**(5):437-448
- [4] Houssein EH, Kilany M, Hassanien AE. Ecg signals classification: A review. *International Journal of Medical Engineering and Informatics*. 2017;**5**(4):376-396
- [5] Banerjee S, Mitra M. Application of cross wavelet transform for ecg pattern analysis and classification. *IEEE Transactions on Instrumentation and Measurement*. 2014;**63**(2): 326-333
- [6] Kærgaard K, Jensen SH, Puthusserypady S. A comprehensive performance analysis of EEMD-BLMS and DWT-NN hybrid algorithms for ECG denoising. *Biomedical Signal Processing and Control*. 2016;**25**:178-187
- [7] Nazarahari M, Namin SG, Markazi AHD, Anaraki AK. A multi-wavelet optimization approach using similarity measures for electrocardiogram signal classification. *Biomedical Signal Processing and Control*. 2015;**20**:142-151
- [8] Rui R, Couto P. A neural network approach to ECG denoising. *CoRR*, abs/1212.5217,2012
- [9] Xiong P, Wang H, Liu M, Zhou S, Hou Z, Liu X. ECG signal enhancement based on improved denoising auto-encoder. *Engineering Applications of Artificial Intelligence*. 2016;**52**(C):194-202
- [10] Xiong P, Wang H, Liu M, Lin F, Hou Z, Liu X. A stacked contractive denoising auto-encoder for ECG signal denoising. *Physiological Measurement*. 2016;**37**(12):2214
- [11] Zhou L, Yan Y, Qin X, Yuan C, Que D, Wang L. Deep learning-based classification of massive electrocardiography data. In: *Advanced Information Management, Communicates, Electronic and Automation Control Conference*; IEEE; 2017. pp. 780-785
- [12] Yan Y, Qin X, Wu Y, Zhang N, Fan J, Wang L. A Restricted Boltzmann Machine Based Two-Lead Electrocardiography Classification. In: *Wearable and Implantable Body Sensor Networks (BSN)*, IEEE 12th International Conference on IEEE. 2015 Jun 9. pp. 1-9
- [13] Rangayyan, Rangaraj M. *Biomedical Signal Analysis: A Case-Study Approach*. Piscataway, NJ: IEEE Press, 2002
- [14] Qi H, Liu X, Pan C. Discrete Wavelet Soft Threshold Denoise Processing for ECG Signal. In: *International Conference on Intelligent Computation Technology and Automation*; Vol. 2; IEEE Computer Society; 2010. pp. 126-129
- [15] Ziarani AK, Konrad A. A nonlinear adaptive method of elimination of power line interference in ECG signals. *IEEE Transactions on Bio-Medical Engineering*. 2002;**49**(6):540
- [16] Alfaouri M, Daqrouq K. ECG signal denoising by wavelet transform thresholding. *American Journal of Applied Sciences*. 2008;**5**(3):276-281
- [17] Dewangan NK, Kowar MK. A review on ECG signal de-noising, QRS complex, P and T wave detection techniques. *International journal of innovative research in electrical, electronics, instrumentation and control engineering*. 2015;**3**(2):10-14

- [18] El Hanine M, Abdelmounim E, Haddadi R, Belaguid A. Electrocardiogram signal denoising using discrete wavelet transform. In: *Technology of Computer: English Edition*; Vol. 2; 2014. pp. 98-104
- [19] Blanco-Velasco M, Weng B, Barner KE. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine*. 2008;**38**(1):1-13
- [20] Li N, Li P. An improved algorithm based on EMD-wavelet for ECG signal de-noising. In: *International Joint Conference on Computational Sciences and Optimization*; Vol. 1; IEEE; 2009. pp. 825-827
- [21] Chan HL, Chen GU, Lin MA, Fang SC. Heartbeat detection using energy thresholding and template match. In: *International Conference of the Engineering in Medicine and Biology Society, 2005 (IEEE-EMBS 2005)*; Vol. 6; IEEE; 2006. pp. 6668-6670
- [22] Krasteva V, Jekova I. QRS template matching for recognition of ventricular ectopic beats. *Annals of Biomedical Engineering*. 2007;**35**(12):2065
- [23] Zhou Y, Hu X, Tang Z, Ahn AC. Sparse representation-based ECG signal enhancement and QRS detection. *Physiological Measurement*. 2016;**37**(12):2093
- [24] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*. 2000;**101**(23):E215
- [25] Chiu C-C, Lin T-H, Liau B-Y. Using correlation coefficient in ECG waveform for arrhythmia detection. *Biomedical Engineering Applications Basis & Communications*. 2005;**17**(03):0500023
- [26] Rifai S, Vincent P, Muller X, Glorot X, Bengio Y. Contractive auto-encoders: Explicit invariance during feature extraction. In: *Proceedings of the 28th International Conference on International Conference on Machine Learning*. Omnipress; 2011 Jun 28. pp. 833-840
- [27] Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. In: *International Conference on Machine Learning*; ACM; 2008. pp. 1096-1103
- [28] Liu G, Luan Y. An adaptive integrated algorithm for noninvasive fetal ECG separation and noise reduction based on ICA-EEMD-WS. *Medical & Biological Engineering & Computing*. 2015;**53**(11):1113
- [29] Rahhal MMA, Bazi Y, Alhichri H, Alajlan N, Melgani F, Yager RR. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences*. 2016;**345**(C): 340-354
- [30] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*. 2010;**11**(12):3371-3408

- [31] Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Briefings in Bioinformatics*. 2016;**18**(5):851
- [32] Bouthillier X, Konda K, Vincent P, Memisevic R. Dropout as data augmentation. *arXiv preprint arXiv:1506.08700*. 2015 Jun 29
- [33] Carbonetto P, Stephens M. Scalable variational inference for bayesian variable selection in regression, and its accuracy in genetic association studies. *Bayesian Analysis*. 2012; **7**(1):73-107
- [34] Challis E, Barber D. Gaussian Kullback-Leibler Approximate Inference. 2013. Available from: [JMLR.org](http://jmlr.org)
- [35] Foti NJ, Xu J, Laird D, Fox EB. Stochastic variational inference for hidden Markov models. In: *International Conference on Neural Information Processing Systems*; Vol. 4; MIT Press; 2014. pp. 3599-3607
- [36] Kingma DP, Welling M. *Auto-Encoding Variational Bayes*. New York: Springer-Verlag; 2013
- [37] Blei DM, Kucukelbir A, Mcauliffe JD. Variational inference: A review for statisticians. *Journal of the American Statistical Association*. 2017;**112**(518):859-877

