International Journal of Computational Intelligence and Applications
Vol. 15, No. 4 (2016) 1650021 (14 pages)

DOI: 10.1142/S1469026816500218

A Novel Method for Classification of ECG Arrhythmias Using Deep Belief Networks

Zhiyong Wu[*,†,‡], Xiangqian Ding[*,§] and Guangrui Zhang[*,¶]

[*]*College of Information Science and Engineering*
Ocean University of China, No. 23 Hongkong West Road
Qingdao, Shandong 266073, China

[†]*School of Computer Science and Technology*
Shandong University of Technology
No. 266 New Village West Road
Zibo, Shandong 255000, China
[‡]*wuzhiyong_sdut@sina.com*
[§]*dingxq1995@vip.sina.com*
[¶]*zhang.guangrui@outlook.com*

Received 1 June 2016
Revised 24 October 2016
Accepted 2 November 2016
Published 25 November 2016

In this paper, a novel approach based on deep belief networks (DBN) for electrocardiograph (ECG) arrhythmias classification is proposed. The construction process of ECG classification model consists of two steps: features learning for ECG signals and supervised fine-tuning. In order to deeply extract features from continuous ECG signals, two types of restricted Boltzmann machine (RBM) including Gaussian–Bernoulli and Bernoulli–Bernoulli are stacked to form DBN. The parameters of RBM can be learned by two training algorithms such as contrastive divergence and persistent contrastive divergence. A suitable feature representation from the raw ECG data can therefore be extracted in an unsupervised way. In order to enhance the performance of DBN, a fine-tuning process is carried out, which uses backpropagation by adding a softmax regression layer on the top of the resulting hidden representation layer to perform multiclass classification. The method is then validated by experiments on the well-known MIT-BIH arrhythmia database. Considering the real clinical application, the inter-patient heartbeat dataset is divided into two sets and grouped into four classes (N, S, V, F) following the recommendations of AAMI. The experiment results show our approach achieves better performance with less feature learning time than traditional hand-designed methods on the classification of ECG arrhythmias.

Keywords: ECG arrhythmias classification; restricted Boltzmann machine; deep belief networks; deep learning.

1. Introduction

Nowadays, many devices of cardiac event monitoring are used to detect arrhythmia by recording the heartbeat signals. For the typical small dataset of electrocardiograph

‡Corresponding author.

(ECG) signals, cardiologists can diagnose various cardiac arrhythmias by visual detection. However, the efficiency of visual detection is limited for vast ECG data collected by devices such as the Holter and loop recorder, which are continuous ECG which can record 24 h, 48 h and even 14 months heartbeat signals,[1] in this situation, auto-classification techniques must be used to analyze.

The main challenges in ECG classification are feature learning and classification algorithms. Many feature extraction methods of ECG signal have been proposed in the literatures such as morphological features,[2] temporal intervals,[2–4] wavelet transform,[3,5] and statistical features.[6] In order to get the most suitable set of features, multiple feature extraction methods are often combined in applications. Moreover, feature reduction techniques such as principal component analysis[7–9] and independent component analysis[10,11] have also been applied to project hundreds of ECG features into a lower dimensional feature space. Once the lower dimensional feature space is defined, the classification models for arrhythmia can be constructed by using intelligence algorithms such as support vector machines,[12–14] artificial neural networks,[15,16] optimum-path forest[17] and swarm intelligence algorithms,[18,19] etc.

Although the abovementioned ECG classification techniques acquire high accuracy using standard dataset (such as the MIT-BIH dataset), there are several issues: (1) the process of feature extraction requires participation of the ECG data expertise that consumes more time and cost; (2) some feature information in the ECG source data could be lost since the ECG feature extraction method is hand-crafted; (3) the constructed model of ECG classification has low adaptability in the analysis of another patient. In order to solve the abovementioned problems, deep learning (DL) attracted much attention in recent years.[20] The idea of DL is to learn a layer of good feature representations automatically from the input data.[21–23] Compared with traditional methods, DL has shown outstanding results in many applications such as image classification,[21,24,25] speech recognition[26] and Physiological data.[27–29] Typical DL architectures consist of deep belief networks (DBN),[30] stacked auto-encoder (SAE),[31] and convolutional neural networks (CNN).[23] Some researchers have achieved positive results on the ECG classification using DL technology. Kiranyaz *et al.*[32] used 1D CNN for patient-specific ECG real-time classification. Rahhal *et al.*[29] used SAE to learn features from the raw ECG data for ECG classification. Wang and Shang[33] used DBN to automatically extract features from raw unlabeled physiological data.

In this paper, a new method based on DBN for ECG arrhythmias classification is proposed. The DBN is stacked with two types of restricted Boltzmann machine (RBM) such as Gaussian–Bernoulli (GBRBM) and Bernoulli–Bernoulli (BBRBM), it can automatically learn features from raw ECG signals without expert interaction. Two training algorithms including contrastive divergence (CD) and persistent contrastive divergence (PCD) are used to adjust the RBM parameters. In order to construct this classifier, a softmax regression layer is added on the top of DBN to perform multiclass classification and supervised fine-tuning. Then the proposed approach is validated experimentally by tests on real ECG signals from the well-known

MIT-BIH arrhythmia database[34] and following the recommendations of association for the advancement of medical instrumentation (AAMI).

The rest of the paper is organized as follows: Sec. 2 presents the classification model of the ECG arrhythmias based on DBN, the theory of RBM and training algorithms. The experiments and results obtained from the experiments on MIT-BIH dataset are discussed in Sec. 3. Finally, conclusions and acknowledgements are drawn in Secs. 4 and 5.

2. Proposed Method

The construction of ECG arrhythmias classification system normally includes four steps: (1) preprocessing; (2) heartbeat segmentation; (3) feature learning; and (4) classification. The four steps are interrelated, each of which determines the quality of classification result. The first two steps of preprocessing and heartbeat segmentation are similar to other traditional methods, which aim to gain filtered heartbeat dataset. In the step of preprocessing all artifact signals such as baseline wander, power line interference and high-frequency noise are removed from the ECG signal. Then some QRS-complex detection methods are used to locate all heartbeats for segmentation. The feature learning stage is the key to success in ECG arrhythmias classification. An unsupervised feature learning method is proposed based on DBN, a softmax regression model is applied to perform multiclass classification and supervised fine-tuning. In the following subsections, DBN architecture for ECG signal learning is firstly introduced. Then, after describing the theoretical background of RBMs (BBRBM and GBRBM), the classification model based on the proposed DBN is explained. Flowchart of the proposed method is shown in Fig. 1.

2.1. *Feature learning using DBN*

The DBN was first proposed by Hinton in 2006, who states a DL architecture can model high-level abstractions in data by multiple nonlinear transformations. Its

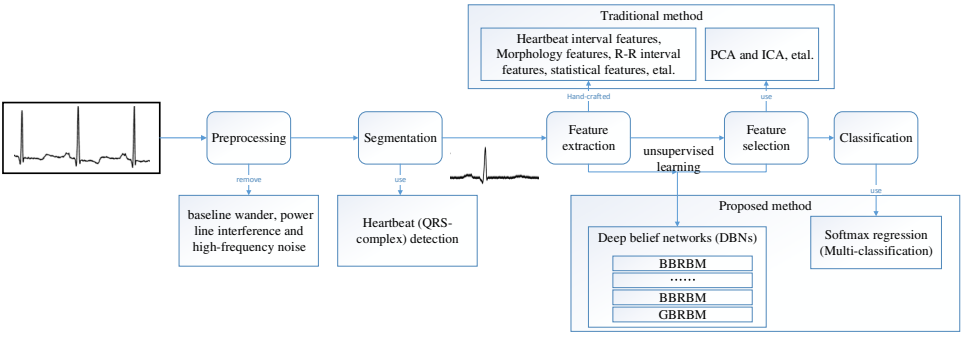

Fig. 1. Flowchart of the proposed method.

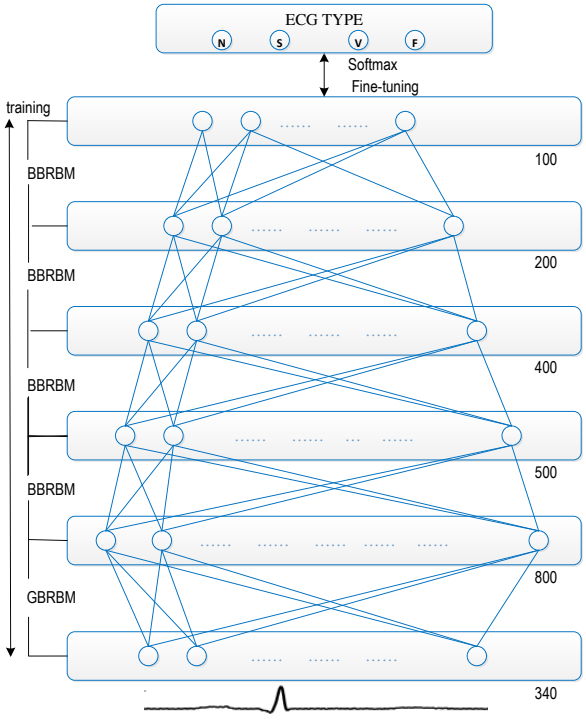

Fig. 2. A DBN with six layers for ECG learning.

highest level output representation can be used as an input to a standalone supervised classification algorithm. Actually, a DBN is structured with stacked RBMs, DBN consists of one visible layer and multiple hidden layers. In this paper, a six-layer DBN with one visible layer and five hidden layers is proposed as shown in Fig. 2. For learning the continuous ECG data, a GBRBM is used to accept the raw heartbeat data at the bottom of the DBN. In general, the build process of the DBN includes two steps: unsupervised training and supervised fine-tune. The unsupervised training of a DBN is completed through sequential training of each individual RBM structure using the RBM learning rule such as CD and PCD, and which will be presented in the next subsection. Then the supervised fine-tune aims to optimize the parameters of the DBN with labeled ECG data.

2.2. *DBN training*

2.2.1. *The theory of RBM*

Boltzmann machine (BM) was first proposed by Hiton and Sejnowski in 1980s.[35–37] A BM as stochastic neural network model can be used to unsupervised learn important aspects of an unknown probability distribution based on samples, while the learning process is difficult and time consuming. In order to overcome these problems, Smolensky proposed a restricted BM (RBM) in 1986.[38] The RBM is similar to

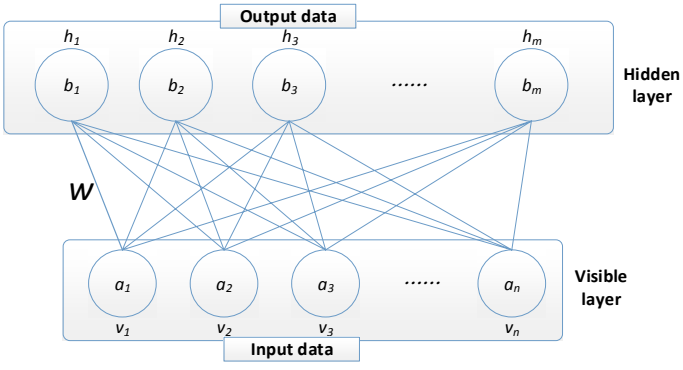

Fig. 3. The undirected graph of RBM with m hidden and n visible variables.

the classical BM, and it can be represented as a bipartite undirected graphical model as shown in Fig. 3.

A RBM consists of one layer of visible units with input $V = (v_1, v_2, \ldots, v_n)$, one layer of hidden units with output $H = (h_1, h_2, \ldots, h_m)$, the undirected weights denoted by w between the visible layer and the hidden layer, two biases units represented by vectors a and b. All visible units are connected to all hidden units without visible–visible or hidden–hidden connections. According to the value types of visible and hidden units, RBM can be divided into BBRBM and GBRBM.[39–41] The two RBMs have the binary-value output in the hidden units, while in the visible units the BBRBM is designed using only binary-value input ($v\epsilon\ \{0,1\}^i$) and the GBRBM is designed using continuous-valued input ($v\epsilon R^i$).

A probability distribution over the joint (v, h) of the visible and the hidden units has an energy function. For BBRBM and GBRBM, the energy functions can be defined in Eq. (1) and Eqs. (2), respectively.[41]

$$E(v, h; \theta_1) = -\sum_{i=1}^{n} v_i a_i - \sum_{j=1}^{m} h_j b_j - \sum_{i=1}^{n} \sum_{j=1}^{m} v_i h_j w_{ij}, \tag{1}$$

$$E(v, h; \theta_2) = \sum_{i=1}^{n} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{v_i}{\sigma_i} h_j w_{ij}, \tag{2}$$

where v_i and h_j are the data states of input data and output data, o'$_i$ being the standard deviation of the Gaussian noise for visible unit $i, \Theta_1 = \{w_{ij}, a_i, b_j\}$ and $\Theta_2 = \{w_{ij}, a_i, b_j, \sigma_i\}$ being the model parameters. Then the joint distribution of v and h is defined in Eq. (3), the marginal distributions of v and h in Eqs. (4) and (5).

$$p(v, h|\theta_{1,2}) = \frac{e^{-E(v,h|\theta_{1,2})}}{\sum_v \sum_h e^{-E(v,h|\theta_{1,2})}}, \tag{3}$$

$$p(v|\theta_{1,2}) = \frac{\sum_h e^{-E(v,h|\theta_{1,2})}}{\sum_v \sum_h e^{-E(v,h|\theta_{1,2})}}, \tag{4}$$

$$p(h|\theta_{1,2}) = \frac{\sum_v e^{-E(v,h|\theta_{1,2})}}{\sum_v \sum_h e^{-E(v,h|\theta_{1,2})}}. \tag{5}$$

Because there are no hidden–hidden or visible–visible connections, for BBRBM, the conditional probabilities of i visible and j hidden units can be defined by Eqs. (6) and (7); for GBRBM, the conditional probabilities of i visible and j hidden units are defined by Eqs. (8) and (9).

$$p(v_i = 1|h, \theta_1) = \delta\left(a_i + \sum_{j=1}^{m} h_j w_{ij}\right), \tag{6}$$

$$p(h_j = 1|v, \theta_1) = \delta\left(b_j + \sum_{i=1}^{n} v_i w_{ij}\right), \tag{7}$$

$$p(v_i = v|h, \theta_2) = \eta\left(v|a_i + \sum_{j=1}^{n} h_j w_{ij}, \sigma_i^2\right), \tag{8}$$

$$p(h_j = 1|v, \theta_2) = \delta\left(b_j + \sum_{i=1}^{n} \frac{v_i}{\sigma_i^2} w_{ij}\right), \tag{9}$$

where the $\delta = 1/(1 + e^{-x})$ is the logistic function, the $\eta(.|u, o')$ denotes the Gaussian probability density function with mean μ and variance o'. In order to easily train the GBRBM using CD algorithm, each component of the input data is normalized to have zero mean and unit variance.[42]

2.2.2. *Training algorithm of RBM*

Given a set of ECG training dataset $V = (v_1, v_2, v_3 \ldots, v_n)$, the goal is to maximize the average log probability of the dataset as shown in Eq. (10).

$$\theta_{1,2}^* = \arg\max \log(\theta_{1,2}) = \arg\max \sum_{i=1}^{n} \log p(v_i|\theta_{1,2}). \tag{10}$$

Then the derivatives of the log probability of a BBRBM training with respect to $\Theta_1 = \{w_{ij}, a_i, b_j\}$ can be calculated as shown in Eq. (11).

$$\begin{aligned} \frac{\partial \log p(v|\theta_1)}{\partial w_{ij}} &= \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}, \\ \frac{\partial \log p(v|\theta_1)}{\partial a_i} &= \langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{model}}, \\ \frac{\partial \log p(v|\theta_1)}{\partial b_j} &= \langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{model}}. \end{aligned} \tag{11}$$

The derivatives of the log probability of a GBRBM training vector with respect to $\Theta_2 = \{w_{ij}, a_i, b_j, o'_i\}$ can be computed as shown in Eq. (12).

$$\begin{aligned}
\frac{\partial \log p(v|\theta_2)}{\partial w_{ij}} &= \left\langle \frac{v_i}{\sigma_i^2} h_j \right\rangle_{\text{data}} - \left\langle \frac{v_i}{\sigma_i^2} h_j \right\rangle_{\text{model}} \\
\frac{\partial \log p(v|\theta_2)}{\partial a_i} &= \left\langle \frac{v_i}{\sigma_i^2} \right\rangle_{\text{data}} - \left\langle \frac{v_i}{\sigma_i^2} \right\rangle_{\text{model}} \\
\frac{\partial \log p(v|\theta_2)}{\partial b_j} &= \langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{model}}
\end{aligned} \tag{12}$$

In Eqs. (11) and (12), data is the probability distribution $P(h|v, \theta_{1,2})$, model is the probability distribution $P(v, h|\theta_{1,2})$. Obviously, data is easy to calculate. In order to compute the intractable $P(v, h|\theta_{1,2})$, CD[43] and PCD[44] have been proposed.

The CD algorithm was first proposed by Hinton in 2002, it has become a standard way to train RBMs. The idea of CD is that a Gibbs sampling (often just one step) is initialized with the training data in the positive phase instead of approximating the negative sample from the RBM-distribution in the log-likelihood gradient. The update rules for parameters of BBRBM and GBRBM can be defined, respectively, as shown Eqs. (13) and (14).

$$\begin{aligned}
\Delta w_{ij} &\approx \varepsilon(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}), \\
\Delta a_i &\approx \varepsilon(\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{model}}), \\
\Delta b_j &\approx \varepsilon(\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{model}}),
\end{aligned} \tag{13}$$

$$\begin{aligned}
\Delta w_{ij} &\approx \varepsilon\left(\left\langle \frac{v_i}{\sigma_i^2} h_j \right\rangle_{\text{data}} - \left\langle \frac{v_i}{\sigma_i^2} h_j \right\rangle_{\text{model}} \right) \\
\Delta a_i &\approx \varepsilon\left(\left\langle \frac{v_i}{\sigma_i^2} \right\rangle_{\text{data}} - \left\langle \frac{v_i}{\sigma_i^2} \right\rangle_{\text{model}} \right) \\
\Delta b_j &\approx \varepsilon\left(\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{model}} \right)
\end{aligned} \tag{14}$$

where ε is learning rate, model denotes the distribution after one step of Gibbs sampling. This method reduces the variance of the gradient estimator and still moves in a direction that pulls the negative chain samples toward the associated positive chain samples.

Based on the CD algorithm, Tieleman proposed a PCD algorithm in 2008. At each gradient update, the PCD algorithm initializes the chain at the last state of the chain used for the previous update, not like the CD initializing the Gibbs chain at the positive phase samples. Although all model parameters are changed in each step, that can receive good samples from model distribution with a few Gibbs sampling steps because the model parameters change slightly. A problem of PCD is how to get the well mixing rate. Tieleman and Hinton proposed quite a different approach of fast PCD (FPCD) in 2009 to reach a faster mixing of the Gibbs chain by introducing additional parameters.[45]

2.3. *DBN fine-tuning*

After layer-by-layer pre-training of DBN, a softmax regression layer can be added on top of the resulting hidden representation layers to perform classification. Softmax regression is a supervised and multi-class learning algorithm. Given training set

$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\}$ of m samples, where $y^{(i)} \epsilon \{1, 2, \ldots k\}$, the probability matrix $h_\Theta(x^{(i)})$ can be computed as Eq. (15).

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|x^{(i)}; \theta) \\ p(y^{(i)} = 1|x^{(i)}; \theta) \\ \ldots \\ p(y^{(i)} = 1|x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \ldots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}, \tag{15}$$

where $\Theta = \{\Theta_1, \Theta_2, \ldots \Theta_k\}$ are the parameters of the model and $1/\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}$ normalizes the distribution. Then the parameters for the entire DBN can be tuned using backpropagation by minimizing the following cost function in Eq. (16).

$$J(\theta) = -\frac{1}{M}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n} \theta_{ij}^2. \tag{16}$$

Many optimization algorithms such as gradient descent and L-BFGS are guaranteed to converge to the global minimum on the basis of formula (16). In this paper, gradient descent is applied to get the gradient variation and cost function minimum.

3. Experimental and Results

3.1. *Data process*

In this study, ECG signals are taken from the MIT-BIH Arrhythmia database, which is developed by Massachusetts Institute of Technology, the ECG recordings were obtained by Beth Israel Hospital Arrhythmia Laboratory.[34] The MIT-BIH Arrhythmia database contains 48 half-hour recordings, sampled at 360 Hz, and 18 types of heartbeats are classified and labeled. In the 48 recordings, 23 recordings are intended to serve as a representative sample of routine clinical recordings and 25 recordings contain complex ventricular, junctional, and supraventricular arrhythmias.

For comparative analysis, the classification system follows the AAMI standardization.[46] This standardization is specified in ANSI/AAMI EC57:1998/(R) 2008 and defines the protocol to perform the evaluations to make sure the experiments are reproducible and comparable. According to the AAMI recommendations, the heartbeat types can be group into five heartbeats classes: (1) normal beat (N), (2) supraventricular ectopic beat (SVEB, here just S), (3) ventricular ectopic beat (VEB, here just V), (4) fusion (F) of a V and a N, and (5) unknown beat type (Q), as shown in Table 1.

To begin the experiment, all ECG signals were filtered with two median filters to remove the baseline wander and a 12-tap low-pass FIR filter with 3 dB point at 35 Hz to remove power-line and high-frequency noise as shown in Ref. 2. Then we utilized the heartbeat fiducial point times provided with the MIT-BIH arrhythmia database to local R peak of the QRS-complex. In the ECG segmentation step, every heartbeat sample had 340 points data including 139 points in front of the R peak and 200 points after the R peak. All samples information of the segmented ECG signals are shown in

Table 1. Mapping the MIT-BIH Arrhythmia types to the AAMI classes.

The AAMI Heartbeat Class Description	N	S	V	F	Q
MIT-BIH heartbeat types (code)	Normal beat (N) Left bundle branch block beat (L) Right bundle branch block beat (R) Atrial escape beat (e) Nodal (junctional) escape beat (j)	Atrial premature beat (A) Aberrated atrial premature beat (a) Nodal (junctional) premature beat (J) Supraventricular premature beat (S)	Premature ventricular contraction (V) Ventricular escape beat (E)	Fusion of ventricular and normal beat (F)	Paced beat (P) Fusion of paced and normal beat (f) Unclassified beat (U)

Table 2. Class Q (unclassified) is discarded since it is marginally represented and has low quantity.

Considering the real clinical application, the heartbeat dataset is defined by inter-patient which has been used in many literatures. According the AAMI recommendation, the four recordings containing paced beats (102, 104, 107, and 217) will be removed from the analysis. Then the left 44 available recordings are divided in two

Table 2. Inter-patient heartbeat dataset of DS1 (training) and DS2 (test).

DS1	N	S	V	F	Total	DS2	N	S	V	F	Total
101	1859	3				100	2237	33			
106	1507		520			103	2080	2			
108	1731	4	17	2		105	2526		41		
109	2489		38	2		111	2115		1		
112	2535	2				113	1788	6			
114	1793	12	43	4		117	1533	1			
115	1952					121	1859	1	1		
116	2287	1	108			123	1514		3		
118	2165	96	16			200	1742	30	826	2	
119	1543		444			202	2060	54	18	1	
122	2474					210	2421	22	193	8	
124	1535	31	46	5		212	2747				
201	1633	121	198	2		213	2640	28	220	362	
203	2527	2	444	1		214	1999		256	1	
205	2568	3	71	11		219	2082	7	64	1	
207	1541	107	193			221	2030		396		
208	1579	2	990	370		222	2268	209			
209	2620	383	1			228	1687	3	362		
215	3194	3	164	1		231	1567	1	2		
220	1952	94				232	398	1382			
223	2044	73	473	14		233	2229	7	830	11	
230	2253		1			234	2700	50	3		
Total	45781	937	3767	412	50897		44222	1836	3216	386	49660

independent datasets as shown in Table 2. The first dataset (DS1: 101, 106, 108, 109, 112, 114, 115, 116, 118,119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, and 230) is the training set, and is used to construct the classification model. The second dataset (DS2: 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, and 234) is the test set, and is used to evaluate the constructed model.

3.2. *Performance evaluation of ECG training*

In the experiment, four different DBNs are designed to extract high-level features and recognize ECG, common training parameters are defined as learning moment ([0.5 0.4 0.3 0.2 0.1 0]), batch size (100), first epoch (50), fine-tuning epoch (200), penalty (0.0002). Two algorithms of CD and PCD are used to train the DBN. The total error accuracy is listed in Table 3. It should be noted that the experiments are carried out on a desktop with the following characteristics (Intel Core i7-4790, CPU 3.6 GHz, RAM 16 GB, and GPU Intel HD graphics 4600). Figure 4 shows the DBN has lower error rate after fine-tuning, and the PCD is not absolutely better than the CD in the ECG data training.

3.3. *Performance evaluation of ECG classification*

In order to evaluate performance, we selected DBN-3, which has the lowest error rate, for ECG classification. The classification performance is measured using the three standard metrics: sensitivity (Se), specificity (Sp) and classification accuracy (Acc), the definitions are shown in Eqs. (17)–(19), respectively.

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{17}$$

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}, \tag{18}$$

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}, \tag{19}$$

Table 3. Error rate of four DBN with CD and PCD algorithms.

DBN		Training Algorithms	Error Rate	
			Before Fine-Tuning	After Fine-Tuning
340-500-200-100	DBN-1	CD	0.0707	0.0513
		PCD	0.0686	0.0511
340-500-200-100-50	DBN-2	CD	0.0769	0.0512
		PCD	0.0751	0.0640
340-600-500-400-200-100	DBN-3	CD	0.0687	0.0652
		PCD	0.0680	0.0511
340-800-500-400-200-100	DBN-4	CD	0.0695	0.0510
		PCD	0.0698	0.0513

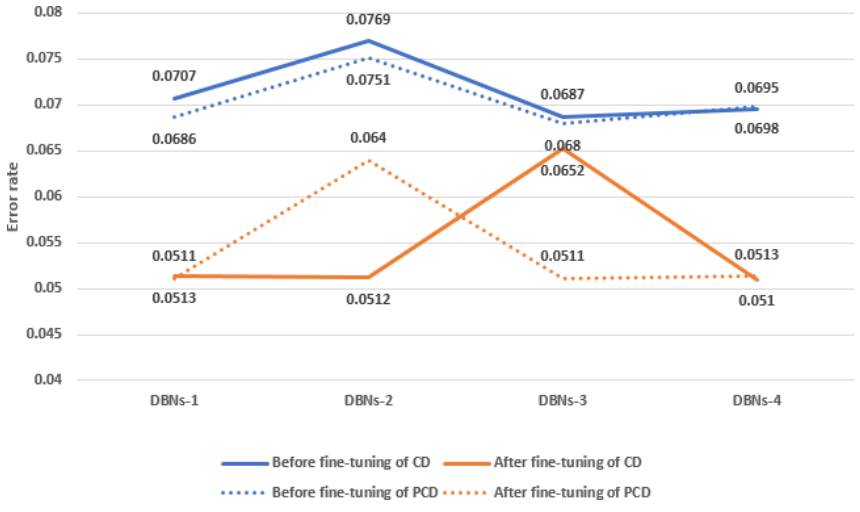

Fig. 4. Line chart of error rate for DBN training.

Table 4. Confusion matrix for ECG arrhythmias classification on DS2 using the DBN-3+softmax.

	Recognized				
Heartbeat Class	N	S	V	F	Total
N	43571	138	477	36	44222
S	275	1556	5	0	1836
V	538	2	2590	86	3216
F	61	6	110	209	386

where TP (true positives) stand for the number of heartbeats of a given class correctly classified, FN (false negatives) stand for the number of heartbeats of a given class incorrectly classified, TN (true negatives) stands for number of the heartbeats not belonging to a given class classified, FP (false positives) stands for the number of heartbeats incorrectly classified as belonging to a given class. Table 4 shows the confusion matrix for ECG arrhythmias classification on DS2 using the DBN-3+softmax.

Table 5. Classification results in terms of V, S, and all using DS2 of MIT-BIH.

	S (%)			V (%)		
Method	Se	Sp	Acc	Se	Sp	Acc
de Chazal *et al.*[2]	75.9	N/A	94.6	77.5	N/A	96.4
Chazal and Reilly[47]	87.7	N/A	95.9	94.3	N/A	99.4
Ince *et al.*[48]	81.8	98.5	96.1	90.3	98.8	97.9
Jiang *et al.*[49]	74.9	98.8	97.5	94.3	99.4	98.8
Proposed	90.2	99.7	99.3	85.8	98.8	97.9

Performance comparison is presented in terms of S and V as shown in Table 5. From Table 5, the proposed method has better accuracy, especially the S class (Se 90.2%, Sp 99.7%, and Acc 99.3%).

4. Conclusion

In this paper, a novel approach for classification of ECG arrhythmias using DBN is proposed. Compared to handcrafted feature-based traditional methods, the proposed approach has several desirable proprieties: (1) it automatically learns abstract feature representations from the raw ECG using DBN and (2) the constructed classification model, which is tested with inter-patient data, have high adaptability for different patients. The experimental results obtained from the tests on MIT-BIH Arrhythmia data achieve better accuracy than traditional hand-designed methods on the classification of ECG arrhythmias. Future researches will focus on increasing the classification accuracy and DL-based dynamic modeling method.

Acknowledgments

This study is partially supported by The National Key Technology R&D Program of China (Grant No. 2015BAF04B02), the Funds for strategic development planning project of Qingdao (Grant No. 14-8-1-7-gx) and health services industry clusters oriented technology service innovation pilot project in 2015.

References

1. J. E. P. Waktare and M. Malik, Holter, loop recorder, and event counter capabilities of implanted devices, *Pac. Clin. Electrophys.* **20** (1997) 2658–2669.
2. P. de Chazal, M. O'Dwyer and R. B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, *IEEE Trans. Biomed. Eng.* **51** (2004) 1196–1206.
3. O. T. Inan, L. Giovangrandi and G. T. A. Kovacs, Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features, *IEEE. Trans. Biomed. Eng.* **53** (2006) 2507–2515.
4. M. L. Soria and J. Martinez, Analysis of multidomain features for ECG classification, in *Computers in Cardiology Conf.*, 2009, pp. 561–564.
5. R. J. Martis, U. R. Acharya and L. C. Min, ECG beat classification using PCA, LDA, ICA and discrete wavelet transform, *Biomed Signal Process Control.* **8** (2013) 437–448.
6. A. Kampouraki, G. Manis and C. Nikou, Heartbeat time series classification with support vector machines, *IEEE Trans. Inf. Technol. Biomed.* **13** (2009) 512–518.
7. K. Polat and S. Güneş, Detection of ECG arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine, *Appl. Math. Comput.* **186** (2007) 898–906.
8. R. J. Martis, U. R. Acharya, K. M. Mandana, A. K. Ray and C. Chakraborty, Application of principal component analysis to ECG signals for automated diagnosis of cardiac health, *Expert. Syst. Appl.* **39** (2012) 11792–11800.

9. R. Rodríguez, A. Mexicano, J. Bila, S. Cervantes and R. Ponce, Feature extraction of electrocardiogram signals by applying adaptive threshold and principal component analysis, *J. Appl. Res. Tech.* **13** (2015) 261–269.
10. S.-N. Yu and K.-T. Chou, Integration of independent component analysis and neural networks for ECG beat classification, *Expert Syst. Appl.* **34** (2008) 2841–2846.
11. S.-N. Yu and K.-T. Chou, Selection of significant independent components for ECG beat classification, *Expert Syst. Appl.* **36** (2009) 2088–2096.
12. E. D. Übeyli, ECG beats classification using multiclass support vector machines with error correcting output codes, *Digit. Signal Process.* **17** (2007) 675–684.
13. C.-P. Shen, W.-C. Kao, Y.-Y. Yang, M.-C. Hsu, Y.-T. Wu and F. Lai, Detection of cardiac arrhythmia in electrocardiograms using adaptive feature extraction and modified support vector machines, *Expert Syst. Appl.* **39** (2012) 7845–7852.
14. A. F. Khalaf, M. I. Owis and I. A. Yassine, A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines, *Expert Syst. Appl.* **42** (2015) 8361–8368.
15. R. Ceylan and Y. Özbay, Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network, *Expert Syst. Appl.* **33** (2007) 286–295.
16. S. Shadmand and B. Mashoufi, A new personalized ECG signal classification algorithm using block-based neural network and particle swarm optimization, *Biomed Signal Process. Control.* **25** (2016) 12–23.
17. E. J. d. S. Luz, T. M. Nunes, V. H. C. de Albuquerque, J. P. Papa and D. Menotti, ECG arrhythmia classification based on optimum-path forest, *Expert Syst. Appl.* **40** (2013) 3561–3573.
18. M. Korürek and A. Nizam, A new arrhythmia clustering technique based on ant colony optimization, *J. Biomed. Inf.* **41** (2008) 874–881.
19. S. Dilmac and M. Korurek, ECG heart beat classification method based on modified ABC algorithm, *Appl. Soft. Comput.* **36** (2015) 641–655.
20. Y. Bengio, A. Courville and P. Vincent, Representation learning: A review and new perspectives, *IEEE. Trans. Pattern. Anal. Mach. Intell.* **35** (2013) 1798–1828.
21. G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* **313** (2006) 504–507.
22. M. Längkvist, L. Karlsson and A. Loutfi, A review of unsupervised feature learning and deep learning for time-series modeling, *Pattern. Recognit. Lett.* **42** (2014) 11–24.
23. J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Netw.* **61** (2015) 85–117.
24. H. C. Shin, M. R. Orton, D. J. Collins, S. J. Doran and M. O. Leach, Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data, *IEEE. Trans. Pattern. Anal. Mach. Intell.* **35** (2013) 1930–1943.
25. M. Hayat, M. Bennamoun and S. An, Deep reconstruction models for image set classification, *IEEE Trans. Pattern. Anal. Mach. Intell.* **37** (2015) 713–727.
26. A. Graves, A. R. Mohamed and G. Hinton, Speech recognition with deep recurrent neural networks, in *2013 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, (2013), pp. 6645–6649.
27. H. Cecotti and A. Graeser, Convolutional neural network with embedded Fourier transform for EEG classification, in *19th Int. Conf. Pattern Recognition*, (2008), pp. 1–4.
28. L. K. M. Längkvist and A. Loutfi, Sleep stage classification using unsupervised feature learning, *Adv. Artif. Neural Syst.* **2012** (2012) 1–9.
29. M. M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani and R. R. Yager, Deep learning approach for active classification of electrocardiogram signals, *Inf. Sci.* **345** (2016) 340–354.

30. G. E. Hinton, S. Osindero and Y.-W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.* **18** (2006) 1527–1554.
31. Y. Wang, H. Yao and S. Zhao, Auto-encoder based dimensionality reduction, *Neurocomputing.* **184** (2016) 232–242.
32. S. Kiranyaz, T. Ince and M. Gabbouj, Real-time patient-specific ECG classification by 1-D convolutional neural networks, *IEEE Trans. Biomed. Eng.* **63** (2016) 664–675.
33. D. Wang and Y. Shang, modeling physiological data with deep belief networks, *Int. J. Inf. Educ. Technol.* **3** (2012) 505–511.
34. G. B. Moody and R. G. Mark, The impact of the MIT-BIH arrhythmia database, *IEEE Eng. Med. Biol.* **20** (2001) 45–50.
35. G. E. Hinton and T. J. Sejnowski, Optimal perceptual inference, in *Proc. IEEE Conf. 223 on Computer Vision and Pattern Recognition* (Washington DC, 1983), pp. 448–453.
36. G. E. Hinton, T. J. Sejnowski and D. H. Ackley, *Boltzmann Machines: Constraint Satisfaction Networks that Learn* (Carnegie-Mellon University, USA, 1984).
37. D. H. Ackley, G. E. Hinton and T. J. Sejnowski, A learning algorithm for Boltzmann machines, *Cogn. Sci.* **9** (1985) 147–169.
38. P. Smolensky, Information processing in dynamical systems: Foundations of harmony theory, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, Vol. 1 (MIT Press, MA, USA, 1986), pp. 194–281.
39. B. Schölkopf, J. Platt and T. Hofmann, in *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference* (MIT Press, MA, USA, 2007), pp. 153–160.
40. A. Krizhevsky, Learning multiple layers of features from tiny images (University of Toronto, Ontario, 2009) (in Canada).
41. K. Cho, A. Ilin and T. Raiko, in *Artificial Neural Networks and Machine Learning – ICANN 2011: 21st Int. Conf. Artificial Neural Networks, Espoo, Finland, June 14–17, 2011, Proceedings, Part I,* eds. T. Honkela, W. Duch, M. Girolami and S. Kaski (Springer, Berlin, Heidelberg, 2011), pp. 10–17.
42. G. E. Hinton, in *Neural Networks: Tricks of the Trade*, 2nd edn., eds. G. Montavon, G. B. Orr and K.-R. Müller (Springer, Berlin, Heidelberg, 2012), pp. 599–619.
43. G. E. Hinton, Training products of experts by minimizing contrastive divergence, *Neural Comput.* **14** (2002) 1771–1800.
44. T. Tieleman, Training restricted Boltzmann machines using approximations to the likelihood gradient, in *Proc. 25th Int. Conf. Machine Learning* (Helsinki, Finland, 2008), pp. 1064–1071.
45. T. Tieleman and G. Hinton, Using fast weights to improve persistent contrastive divergence, in *Proc. 26th Annual Int. Conf. Machine Learning* (Montreal, Quebec, Canada, 2009), pp. 1033–1040.
46. ANSI/AAMI/ISO EC57, 1998(R) 2008, Testing and Reporting Performance Results of Cardiac Rhythm and ST-Segment Measurement Algorithms (American National Standard, 1999).
47. P. de Chazal and R. B. Reilly, A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features, *IEEE Trans. Biomed. Eng.* **53** (2006) 2535–2543.
48. T. Ince, S. Kiranyaz and M. Gabbouj, A generic and robust system for automated patient-specific classification of ECG signals, *IEEE Trans. Biomed. Eng.* **56** (2009) 1415–1426.
49. W. Jiang and S. G. Kong, Block-based neural networks for personalized ECG signal classification, *IEEE Trans. Neural. Netw.* **18** (2007) 1750–1761.