This is the Pre-Published Version.

The following publication S. S. Xu, M. -W. Mak and C. -C. Cheung, "I-Vector-Based Patient Adaptation of Deep Neural Networks for Automatic Heartbeat Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 3, pp. 717-727, March 2020 is available at https://doi.org/10.1109/JBHI.2019.2919732.

I-Vector Based Patient Adaptation of Deep Neural Networks for Automatic Heartbeat Classification

Sean Shensheng Xu, *Student Member*, Man-Wai Mak, *Senior Member*, and Chi-Chung Cheung, *Senior Member*

***Abstract*—Automatic classification of electrocardiogram (ECG) signals is important for diagnosing heart arrhythmias. A big challenge in automatic ECG classification is the variation in the waveforms and characteristics of ECG signals among different patients. To address this issue, this paper proposes adapting a patient-independent deep neural network (DNN) using the information in the patient-dependent identity vectors (i-vectors). The adapted networks, namely i-vector adapted patient-specific DNNs (iAP-DNNs), are tuned towards the ECG characteristics of individual patients. For each patient, his/her ECG waveforms are compressed into an i-vector using a factor analysis model. Then, this i-vector is injected into the middle hidden layer of the patient-independent DNN. Stochastic gradient descent is then applied to fine-tune the whole network to form a patient-specific classifier. As a result, the adaptation makes use of not only the raw ECG waveforms from the specific patient but also the compact representation of his/her ECG characteristics through the i-vector. Analysis on the hidden-layer activations shows that by leveraging the information in the i-vectors, the iAP-DNNs are more capable of discriminating normal heartbeats against arrhythmic heartbeats than the networks that use the patient-specific ECG only for the adaptation. Experimental results based on the MIT-BIH database suggest that the iAP-DNNs perform better than existing patient-specific classifiers in terms of various performance measures. In particular, the sensitivity and specificity of the existing methods are all under the receiver operating characteristic curves of the iAP-DNNs.**

***Index Terms*—ECG classification; Arrhythmias; Deep neural networks; i-vectors; DNN adaptation.**

I. Introduction

Heart arrhythmias or arrhythmias refer to the irregular heartbeats of patients. Not all arrhythmias are serious or life threatening but some types (e.g., atrial fibrillation, ventricular escape and ventricular fibrillation) may be a sign of heart diseases or even cause sudden cardiac death if prompt treatments are not received.

Arrhythmias can be detected through electrocardiography (ECG), which is a process of recording the electrical activities of the heart. The standard ECG uses a 12-lead configuration in which a number of electrodes are attached to the skin of a patient, and measurements can only be made for a couple of minutes. The conventional 12-lead ECG is therefore not practical for long-term monitoring. Unfortunately, some intermittent arrhythmias can only be detected by long-term monitoring because they can be easily missed in ordinary recording sessions. To overcome this problem, the 2-lead configuration commonly used in the Holter monitor [1] can be used for heart monitoring for up to 48 hours. In this work, we focus on the long-term continuous cardiac monitoring. Therefore, we used the MIT-BIH arrhythmia database [2] for performance evaluation, which comprises a standard set of Holter recordings for evaluating arrhythmia detectors.

With the increasing use of personal portable devices to acquire ECG data, a large number of ECG recordings can be collected. However, it is impossible to read and analyze all of these data manually by medical professionals. It is better to use machines to classify heartbeats automatically so as to assist clinicians in diagnosing arrhythmias. One of the biggest challenges in automatic heartbeat classification is the variations in ECG characteristics among different patients, which is known as inter-patient variations. Patient-independent classifiers that are trained on the ECG of a large number of patients may not perform well on unseen patients [3]. While a patient-specific classifier can be trained using patient-specific data only, the performance will not be good. This is because patient-specific data are usually very limited, which either leads to overfitting in complex classifiers or insufficient capability for simple classifiers. This is why most of the studies (e.g., [4]–[9]) used a small amount of patient-specific data together with some patient-independent data to train patient-specific classifiers.

To address the patient-dependent variability in the ECG signals, we have developed a deep neural network (DNN) based heartbeat classifier [10] that is adaptive to the ECG characteristics of individual patients. The adaptation is achieved by using the i-vector representation [11] of patient-specific ECG as auxiliary information to adjust the weights in the DNN. This paper is an extension of our earlier work in [10]. It provides additional analyses to explain why the i-vectors can help adapt the patient-independent DNN. In particular, new experiments have been performed to investigate the best layer for injecting the i-vectors. Visualizations of the network activities during the course of adaptation are provided to demonstrate the effectiveness of i-vector adaptation. Through these investigations, we are able to explain why this i-vector adaptation can lead to patient-specific classifiers that outperform other state-of-the-art patient-specific classifiers.

This paper is organized as follows. Section II provides a survey of previous patient-specific heartbeat classification methods. Section III explains the concept of i-vectors and presents the details of the proposed i-vector adapted DNNs (iAP-DNNs). Section IV outlines the ECG dataset used in this study and the experimental setting. Section V investigates

The authors are with Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong SAR of China. This work was in part supported by the RGC of Hong Kong, Grant No. PolyU152137/17E.

Manuscript received April 19, 2005; revised August 26, 2015.

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

TABLE I: Summary of the existing studies [4]–[9] and the proposed method. All of these methods (patient-specific classifiers) follow the ANSI/AAMI EC57 standard, adopt the beat-by-beat analysis strategy and the subject-oriented evaluation scheme.

Ref.	Feature	Classifier
Jiang *et al.* (2007) [4]	Hermite transform coefficients	BbNN
Ince *et al.* (2009) [5]	Morphological features	MD PSO
Kiranyaz *et al.* (2016) [6]	Raw ECG downsampling and FFT	CNN
Ye *et al.* (2016) [7]	Morphological and dynamic features	SVM based on multiview learning
Zhai *et al.* (2018) [8]	Dual heartbeat coupling matrix	CNN
Li *et al.* (2018) [9]	Raw ECG	TDCNN
Proposed	Raw ECG and i-vectors	DNN

which hidden layer of the DNN is more appropriate for receiving the i-vector injection. It also compares the performance of the proposed iAP-DNNs against the existing patient-specific classifiers using some standard performance metrics. Finally, Section VI draws conclusions and suggests directions for future work.

II. Related Work

In the recent past, there have been much efforts [4]–[9], [12]–[15] in classifying heartbeats automatically. Many of these studies [4]–[9], [12] used the MIT-BIH arrhythmia database for performance evaluation and followed the standard prepared by the Association for the Advancement of Medical Instrumentation (ANSI/AAMI EC57:1998) [16] for testing and reporting performance. They adopted a "subject-oriented" evaluation scheme [12] and used five minutes of patient-specific ECG data for constructing patient-specific classifiers. Under this scheme, data are divided according to patients instead of heartbeats, which ensures that the training set and the test set comprise different patients. As a result, the performance reported in [4]–[9] is more realistic and is closer to the practical situations. A summary of these studies [4]–[9] as well as our proposed method is shown in Table I.

A. Patient-Specific Models for Heartbeat Classification

Jiang *et al.* [4] proposed using Hermite transform coefficients to approximate the QRS complexes of heartbeats. The coefficients and R-R intervals were used as heartbeat features for classification by an evolvable block-based neural network (BbNN) [17]. In the training stage, both common (totally 142 beats from 20 patients) and patient-specific data (5-minute ECG from each patient) were used for evolving the patient-specific BbNNs. The results suggest that high accuracies can be achieved by using personalized ECG classification. However, lots of critical parameters or thresholds needed to be set empirically in this approach.

In [5], wavelet transform and principal component analysis (PCA) were applied to extract morphological features. The low dimensional morphological feature vectors were combined with temporal features to form the final feature vectors. A multi-dimensional particle swarm optimization (MD PSO) method was proposed, which optimizes neural network based classifiers according to 245 common training beats and a variable number of patient-specific beats. Overall, this method achieves performance that is comparable with the BbNN-based personalized ECG classifier in [4].

In [6], the raw data of each beat were downsampled to 64 or 128 time-points centered on the R-peak, and FFT representations were used as the input to a patient-specific 1-D convolutional neural network (CNN). Each CNN was trained by using 245 representative beats that are common to all patients and 5-minute patient-specific beats. Results show that the CNNs outperform any existing arrhythmia classifiers under the same evaluation protocol.

Ye *et al.* [7] utilized wavelet transform and independent component analysis (ICA) to extract morphological features from segmented heartbeats. Unlike other patient-specific classifiers, the classifiers in [7] can be trained on unlabeled patient-specific data, meaning that no manual intervention is required during training. Specifically, a general classifier was trained on the data extracted from the patients who are similar to the target patient. Then, a patient-specific classifier was trained on a small amount of patient-specific ECG with high-confident labels hypothesized by a multi-view model. The final result was obtained by combining the two classifiers probabilistically. Results shown that the customized models together with automatic adaptation can improve classification performance.

In [8], the beats were transformed into dual-beat coupling matrices, which are used as 2-D inputs to a CNN classifier. The matrices captured both beat morphology and beat-to-beat correlation in ECG. A heartbeat selection procedure was also proposed to select the most representative beats. For each patient, a classifier was trained based on these representative beats and the patient-specific ECG. Results demonstrated that the 2-D CNN-based classifiers were superior to several state-of-the-art detectors.

In [9], a generic convolutional neural network (GCNN) was trained based on the ECG of a general population. The GCNN was then fine-tuned to form a tuned dedicated CNN (TDCNN) using patient-specific ECG. Raw ECG signals were used as the input of the CNN classifiers and the heartbeat segmentation procedure was the same as [6]. To explore the influence of the amount of training samples on the performance of TDCNN, 2-, 3-, 4- and 5-minute patient-specific ECG were used to adapt the GCNN. The results show that more training samples help the TDCNN to achieve higher classification accuracy and the performance was comparable with the existing patient-specific classifiers.

B. Comparison of Adaptation Methods

In general, the amount of ECG data from the general population is much larger than that from individual patients for adapting the classifiers. Therefore, the adapted patient-specific classifiers may be biased towards the patterns in the

general population. To overcome this issue, in [4]–[6], [8], the patient-specific classifiers were trained based on common and patient-specific beats. Specifically, the common heartbeats were randomly sampled from the corresponding classes of the general population in [4]–[6] while an automatic selection method was proposed to select the most representative beats in different classes in [8]. After all, the number of selected common beats was limited to a few hundred only. In [7], [9], instead of using the ECG of the entire population, a subset was selected for training the general classifier. However, reducing the amount of data from the general population is not a desirable way to address the issue because it throws away lots of useful information in the ECG data. Also, the common training data are useful when the patient-specific beats contain a few arrhythmia patterns only [2].

In our adaptation method, all of the ECG data from the general population are used for training a patient-independent DNN as shown in Fig. 3(a). Then, for each patient, an i-vector is extracted from his/her 5-minute ECG data. As shown in Fig. 3(b), to form a patient-specific classifier, the i-vector is used as another input to the middle layer of the patient-independent DNN and the whole network is fine-tuned by backpropagation. The patient-independent and patient-specific DNNs represent general population knowledge and specific personal knowledge, respectively [7]. The advantage of the method is that it can leverage all of the ECG data in the general population but still be able to adapt to the ECG characteristics of individual patients through the patient-specific ECG and the patients' i-vectors.

III. Methodology

This section first outline the i-vector extraction process and explain why i-vectors can be used for representing patient-specific information. Then, we present the proposed iAP-DNNs, specifically, showing the architecture of a patient-independent DNN and describing how to migrate it to a patient-specific DNN. In the patient-specific DNN, we not only make use of patient-specific data but also i-vectors of the patient for patient adaption. Thus, we also introduce the procedure to extract an i-vector from a particular patient and describe how to embed the i-vector into the adaption. Finally, we discuss the advantages of the iAP-DNNs.

A. I-vector Extraction

The idea of i-vectors is based on the factor analysis method that compresses speaker and channel information into a low-dimensional subspace [18]. Inspired by the success of i-vectors in representing speaker information, we applied i-vectors to represent patient-specific information in ECG signals.

Fig. 1 illustrates the procedure of training an i-vector extractor given a set of ECG data from a general population; it also shows the process of extracting an i-vector from an ECG record. First, PCA whitening is applied to reduce the correlation among the time-points in the ECG vectors [19]. Then, the whitened ECG vectors from the general population are used to train a Gaussian mixture model, which we referred to as the universal background model (UBM). The ECG data

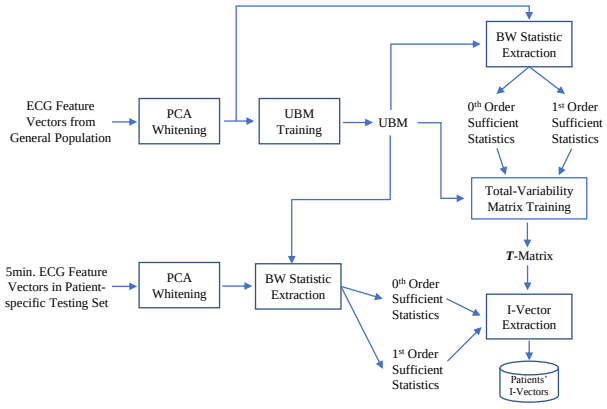


Fig. 1: Training of i-vector extractor and i-vector extraction process.

are then aligned with the UBM to compute the 0th- and 1st-order sufficient statistics (Baum-Welch statistics), from which a total variability matrix (T-matrix) is trained. To extract an i-vector, the same processing pipeline is applied (see the lower branch of Fig. 1) to an ECG record to compute the sufficient statistics. Given the T-matrix and the sufficient statistics, an i-vector representing the whole ECG record can be obtained. In the sequel, we outline the formulae for training an i-vector extractor and the i-vector extraction process. For detailed derivations, readers may refer to [20].

Given the i-th ECG record from a general population, we extract the D-dimensional heartbeat vectors $\mathcal{X}_i = \{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT_i}\}$ from the record, where T_i is the number of complete heartbeats in the record.[1] We assume that the ECG vectors from this record are generated by a C-mixture GMM with parameters $\Lambda_i = \{\pi_c, \boldsymbol{\mu}_{ic}, \boldsymbol{\Sigma}_c\}_{c=1}^{C}$, i.e.,

$$p(\mathbf{x}_{it}) = \sum_{c=1}^{C} \pi_c^{(b)} \mathcal{N}(\mathbf{x}_{it} | \boldsymbol{\mu}_{ic}, \boldsymbol{\Sigma}_c^{(b)}), \qquad t = 1, \ldots, T_i. \quad (1)$$

In Eq. 1, we assume that $\pi_c^{(b)}$ and $\boldsymbol{\Sigma}_c^{(b)}$ are tied across all ECG records and are equal to the mixture weights and covariance matrices of the UBM, respectively.

In the i-vector framework [11], the mean vectors $\{\boldsymbol{\mu}_{ic}\}_{c=1}^{C}$ are stacked to form a GMM-supervector [18] $\boldsymbol{\mu}_i = [\boldsymbol{\mu}_{i1}^{\mathsf{T}} \quad \ldots \quad \boldsymbol{\mu}_{iC}^{\mathsf{T}}]^{\mathsf{T}}$, which is assumed to be generated by the following factor analysis model [21]:

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}^{(b)} + \boldsymbol{T}\boldsymbol{w}_i, \quad (2)$$

where $\boldsymbol{\mu}^{(b)}$ is obtained by stacking the mean vectors of the UBM, $\boldsymbol{T}$ is a $CD \times R$ low-rank total variability matrix modeling all sort of variability in the ECG vectors, and $\boldsymbol{w}_i \in \Re^R$ comprises the latent (total) factors. Eq. 2 suggests that the generated supervectors $\boldsymbol{\mu}_i$'s have mean $\boldsymbol{\mu}^{(b)}$ and covariance matrix $\boldsymbol{T}\boldsymbol{T}^{\mathsf{T}}$. Eq. 2 can also be written in a component-wise form:

$$\boldsymbol{\mu}_{ic} = \boldsymbol{\mu}_c^{(b)} + \boldsymbol{T}_c \boldsymbol{w}_i, \qquad c = 1, \ldots, C \quad (3)$$

where $\boldsymbol{\mu}_{ic} \in \Re^D$ is the c-th sub-vector of $\boldsymbol{\mu}_i$ (similarly for $\boldsymbol{\mu}_c^{(b)}$) and $\boldsymbol{T}_c$ is a $D \times R$ sub-matrix of $\boldsymbol{T}$.

[1] See [19] for the definition of complete heartbeats.

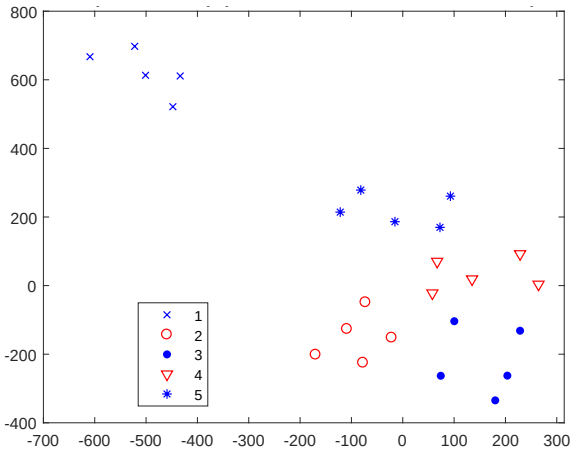


Fig. 2: The i-vectors of five patients projected onto a 2-D t-SNE embedded space. Each patient is represented by one marker and each point represents an i-vector. Patient-dependent clusters are apparent.

In the i-vector framework, every ECG record is assumed to be obtained from a different patient. As a result, the ECG vectors of Record i aligning to mixture c have mean $\boldsymbol{\mu}_{ic}$ and covariance matrix $\boldsymbol{\Sigma}_c^{(b)}$. This matrix measures the deviation of the ECG vectors associated with the c-th mixture from $\boldsymbol{\mu}_{ic}$. In practice, $\boldsymbol{\mu}_c^{(b)}$ and $\boldsymbol{\Sigma}_c^{(b)}$ are the mean vectors and covariance matrices of the UBM. As a result, we only need to estimate the T-matrix $\boldsymbol{T}$ from a set of training ECG vectors.

Assume that there are P ECG recordings from the general population. The T-matrix can be estimated according to the expectation-maximization (EM) algorithm as follows [20]:

- E-step:

$$\langle \boldsymbol{w}_i|\mathcal{X}_i\rangle = \boldsymbol{L}_i^{-1}\sum_{c=1}^{C}\boldsymbol{T}_c^{\mathsf{T}}(\boldsymbol{\Sigma}_c^{(b)})^{-1}\tilde{\boldsymbol{f}}_{ic}, \tag{4}$$

$$\langle \boldsymbol{w}_i\boldsymbol{w}_i^{\mathsf{T}}|\mathcal{X}_i\rangle = \boldsymbol{L}_i^{-1} + \langle \boldsymbol{w}_i|\mathcal{X}_i\rangle\langle \boldsymbol{w}_i|\mathcal{X}_i\rangle^{\mathsf{T}}, \tag{5}$$

$$\boldsymbol{L}_i = \boldsymbol{I} + \sum_{c=1}^{C}\boldsymbol{N}_{ic}\boldsymbol{T}_c^{\mathsf{T}}(\boldsymbol{\Sigma}_c^{(b)})^{-1}\boldsymbol{T}_c; \tag{6}$$

- M-step:

$$\boldsymbol{T}_c = \left[\sum\nolimits_i \tilde{\boldsymbol{f}}_{ic}\langle \boldsymbol{w}_i|\mathcal{X}_i\rangle^{\mathsf{T}}\right]\left[\sum\nolimits_i \boldsymbol{N}_{ic}\langle \boldsymbol{w}_i\boldsymbol{w}_i^{\mathsf{T}}|\mathcal{X}_i\rangle\right]^{-1}, \tag{7}$$

where $i = 1,\ldots,P$, $\langle\cdot|\cdot\rangle$ is the conditional expectation and $\boldsymbol{T}_c$ is the c-th partition of $\boldsymbol{T}$. The 0th-order and the 1st-order Baum-Welch statistics in Eq. 4, Eq. 6 and Eq. 7 can be computed as follows:

$$\begin{aligned}\boldsymbol{N}_{ic} &= \sum\nolimits_t \gamma_c(\mathbf{x}_{it}),\\ \tilde{\boldsymbol{f}}_{ic} &= \sum\nolimits_t \gamma_c(\mathbf{x}_{it})(\mathbf{x}_{it} - \boldsymbol{\mu}_c^{(b)}),\end{aligned} \tag{8}$$

where $\gamma_c(\mathbf{x}_{it})$ is the posterior probability of mixture c.

The i-vector $\mathbf{i}_i \equiv \langle \boldsymbol{w}_i|\mathcal{X}_i\rangle$ representing the i-th ECG recording can be computed according to Eq. 4.

B. I-vector an ECG Representation

Fig. 2 demonstrates why i-vectors are good for representing patient-dependent information, which makes them ideal for adapting ECG classifiers. In the figure, each marker corresponds to one patient and each point of the same marker corresponds to an i-vector extracted from an ECG record of that patient. Totally, there are five patients, each has five ECG records. For ease of visualization, the i-vectors were projected onto an embedding space created by the t-SNE (t-distributed stochastic neighbor embedding) software [22]. T-SNE is a nonlinear dimension reduction method for visualizing high-dimensional data on a two- or three-dimensional space. Apparently, the i-vectors of the same patient are close to each other, i.e., forming patient-specific clusters in the t-SNE space. This clustering phenomenon suggests that the i-vectors can capture patient-specific information, which is very useful for adapting ECG classifiers.

C. Patient-Independent DNN (General Classifier)

Fig. 3(a) shows the architecture a patient-independent ECG classifier. It is essentially a DNN with fixed-length ECG waveforms as the input and heartbeat types as the output. The fixed-length waveforms can be obtained by the segmentation and alignment process described in [19].[2]

To apply DNNs for M-class classification, we can construct a DNN with $L-1$ hidden layers and a softmax output layer with M output nodes. Specifically, denote $a_m^{(L)}$ as the activation of the m-th neuron in the softmax layer, where $m = 1,\ldots,M$, the softmax function gives the outputs:

$$y_m = \frac{\exp\left\{a_m^{(L)}\right\}}{\sum_{j=1}^{M}\exp\left\{a_j^{(L)}\right\}}, \qquad m = 1,\ldots,M. \tag{9}$$

With the softmax function, the outputs can be considered as the posterior probabilities of individual classes given an input vector $\mathbf{x}$, i.e., $y_m \equiv P(\text{Class} = m|\mathbf{x})$. The activation $a_m^{(L)}$ is the linear weighted sum of the hidden nodes' output at the $(L-1)$-th hidden layer.

The patient-independent DNN is trained by the backpropagation algorithm using the ECG data of a number of patients in the general population.

D. Patient-Specific DNN

To create a patient-specific classifier, the weights in the lower part of the general classifier in Fig. 3(a) are retained and the weights in the upper part are randomized. Then, for each patient, five minutes of his/her ECG data are presented to the input and an i-vector extracted from these 5-minute ECG data is injected into the middle layer of the patient-independent DNN, as shown in Fig. 3(b). The whole network is then fine-tuned by backpropagation. The backpropagation algorithm will encourage the upper layers to represent patient-dependent ECG information at a more abstract level. This results in the output layer being tuned to the characteristics of the corresponding patient. The i-vector extracted from the

[2]A GitHub page (https://github.com/seanssx) has been created for other researchers to download the implementation the procedure.

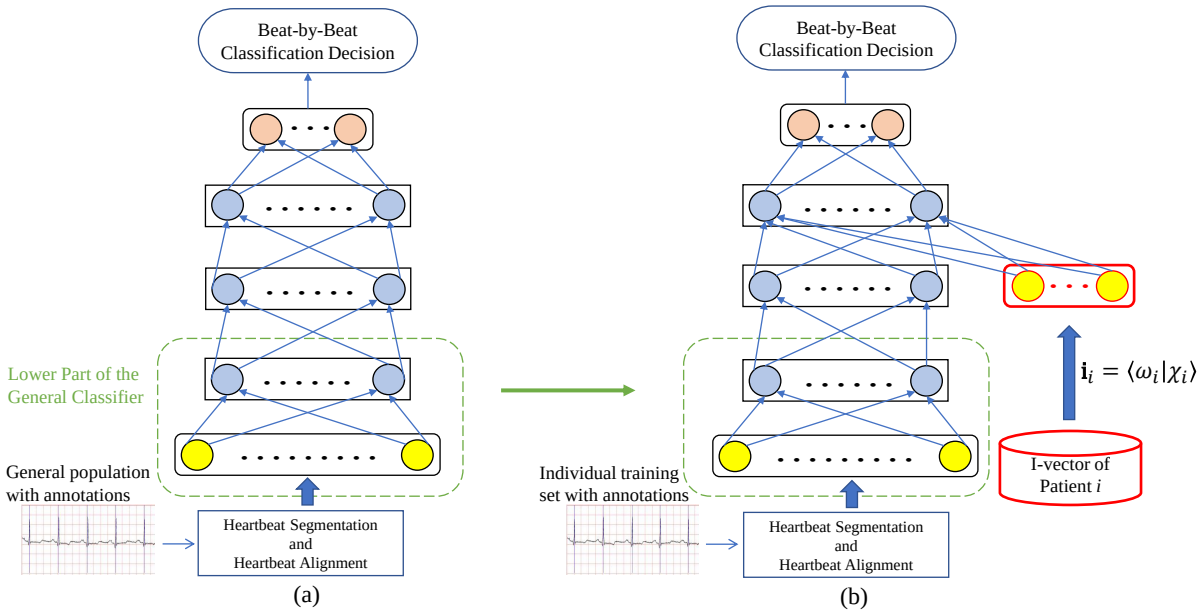


Fig. 3: I-vector adapted patient-specific DNNs (iAP-DNNs). (a) General classier. (b) Patient-specific classifier.

training ECG of a patient is applied to adapt the patient-independent DNN to a patient-dependent DNN. The same i-vector will also be used as an auxiliary input to the adapted DNN (Fig. 3(b)) during testing. This means that the identity of the patient is assumed to be known during testing. But this assumption is reasonable in clinical settings.

The i-vector is presented to the second hidden layer instead of the first hidden layer because it is well known that the feature representation becomes increasingly abstract when moving up the network [23]. For example, in DNN-based speech recognition, the bottom layers can capture low-level acoustic features that vary significantly across different speakers and the upper layers can capture high-level features that are less speaker dependent [24]. This suggests that the upper layer can implicitly normalize the features across speakers. By the same token, the upper layers of the DNN in Fig. 3(a) will produce patient-invariant features, which is not good for patient-specific classification. This explains why it is necessary to use the patient-dependent i-vector to adapt the network. To check the correctness of the above justification, the patient's i-vector was injected into different hidden layers of the network and the results will be shown in Section V-A.

Each patient has a number of heartbeat vectors. Specifically, for the r-th patient, his/her heartbeat vectors are denoted as $\mathcal{X}_r = \{\mathbf{x}_{r1}, \ldots, \mathbf{x}_{rT_r}\}$, where T_r is the number of heartbeats from this patient.[3] On the other hand, each patient has one i-vector only, which is extracted from $\mathcal{X}_r$ using Eq. 4, i.e., $\mathbf{i}_r = \langle \boldsymbol{w}_r | \mathcal{X}_r \rangle$. The backpropagation algorithm, however, requires one input vector for every output vector. To overcome this imbalance in the number of input vectors, we repeated the same i-vector for each ECG vector, as shown in Fig. 4.

Once the DNN has been adapted, it can be used for classifying the ECG of the corresponding patient in a beat-by-beat basis. Specifically, given a test ECG waveform of the patient, its heartbeats are segmented and aligned to form 417-dimensional heartbeat vectors [19]. The heartbeat vectors are presented to the input of the DNN. Meanwhile, the i-vector of this patient is retrieved from the i-vector repository (see Fig. 3(b)). For each heartbeat vector, the i-vector is replicated and presented to the middle layer of the DNN. The outputs of the DNN are then averaged over all of the heartbeat vectors to obtain the posterior probability of individual heartbeat classes.

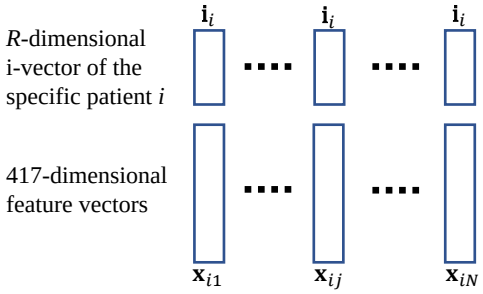


Fig. 4: Repetition of an i-vector to match the number of ECG vectors for each patient. Vectors in top row will be injected into the middle layer of the DNN. Vectors in the bottom row are the input of the DNN.

E. Advantages of iAP-DNNs

To deal with inter-patient variability in ECG signals, existing methods typically use three approaches: (1) pooling the patient-specific and patient-independent data together to train a patient-specific classifier [4]–[6], [8], (2) combining the predictions made by a patient-independent classifier and a patient-specific classifier [7], and (3) fine-tuning a patient-independent classifier using patient-specific data [9]. The major problem of these approaches is that they fail to take advantage of the

[3]In Section III-A, the subscript i refers to ECG recordings. Here, the subscript r refers to patients.

vast amount of ECG signals from the general population. In particular, to prevent the limited amount of patient-dependent data from being overshadowed by the patient-independent data, only a small fraction of the patient-independent data will be used in the first and second approaches. While fine-tuning is a reasonable approach, the information learned from the general population could be easily lost or forgotten if the degree of fine-tuning is substantial.

The iAP-DNNs are designed to overcome the problems in the three approaches mentioned above. The key ideas are (1) to leverage the ECG data of a general population to create a patient-independent DNN and (2) to focus the adaptation on the upper layers of the DNN using patient-specific information to make it patient-dependent. To avoid being overshadowed by the data in the general population, the weights in the upper layers are re-initialized before adaptation begins. To avoid forgetting the learned information from the general population, the bottom layers of the network will only be adapted by a small amount of patient-specific data, i.e., the extent of adaptation in the lower layers will not be substantial. These strategies are superior to the data pooling approach in that it is not necessary to ensure a good balance between the patient-independent and patient-specific data. To gear the adaptation of the upper layers to specific patient, the i-vector that characterizes an individual patient is injected into the middle layer of the network. Results in Section V and Fig. 6 suggest that this step has great impact on the DNN to classify the ECG of individual patients.

Some recent studies [6], [9], [13], [15] applied convolutional neural networks (CNNs) to classify raw ECG signals into different arrhythmia types, primarily because of the intrinsic capability of CNNs in dealing with phase shift variability. In fact, it has been found in speech recognition research that applying max-pooling in time could produce representations that are less sensitive to phase shifts [25]. Our proposed method uses heartbeat segmentation and heartbeat alignment [19] to minimize the phase shift variation, which enables us to use DNN instead of CNN to classify the heartbeats. The question is "Which is a better way to deal with phase shifts: max-pooling or heartbeat alignment?". The answer lies in whether we can detect the R peaks accurately. If we can, heartbeat alignment is a better choice. On the other hand, if aligning the heartbeats is difficult, CNN is a better choice. For the MIT-BIH arrhythmia dataset, heartbeat alignment is a better choice because the R peaks in this dataset can be predicted at an accuracy of over 99% by using the Pan-Tompkins algorithm [26]. In fact, the results in Section V-C also suggest that heartbeat alignment together with the proposed DNN adaptation outperform state-of-the-art CNNs in this dataset.

Another advantage of heartbeat alignment is that DNNs are more amenable to adaptation by i-vectors than CNNs. This is because for ECG classification, the convolutional layers and max-pooling layers of a CNN have the concept of time, which are not compatible with the static information encoded in the i-vectors. Because the hidden layers in a DNN are static, injecting an i-vector into its hidden layers can be considered as shifting the activations of the hidden layers, where the shift accounts for the patient-specific information.

TABLE II: Mapping from MIT-BIH heartbeat types to AAMI heartbeat classes.

AAMI class	MIT-BIH class code	No. of beats
N	NOR, LBBB, RBBB, AE, NE	90042
S	AP, aAP, NP, SP	2779
V	PVC, VE	7007
F	fVN	802
Q	P, fPN, U	15

IV. Experimental Setting

This section first introduces the data set and evaluation protocol in our experiments. Then, we describe some issues concerning the implementation (i.e., DNN structure and DNN training).

A. Data Set

The MIT-BIH arrhythmia database [2] was used for performance evaluation. The database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings of 47 patients. Each record contains a continuous recording of raw ECG signals, which were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. The database provides annotation for both beat-by-beat class information and corresponding time series information (e.g., positions of R peaks) that were verified by two or more cardiologists independently. The total number of labelled heartbeats is 108,655 and these heartbeats are classified into 15 different types. According to the American National Standard prepared by the Association for the Advancement of Medical Instrumentation (ANSI/AAMI EC57:1998), we combined the 15 heartbeat types into five classes (see Table II). They are normal sinus beats (N), supraventricular ectopic beats (S), ventricular ectopic beats (V), fusion of a normal and a ventricular ectopic beat (F) and unknown beat type (Q).

B. Evaluation Protocol

As suggested by the ANSI/AAMI EC57 standard [16], we focused on evaluating the classification performance of two majority arrhythmia classes (Class S and Class V). Besides, four ECG recordings (Record IDs 102, 104, 107 and 217), which contain paced beats, were excluded. As a result, a total of 44 recordings were used for performance evaluation.

We have conducted two experiments (Exp. 1 and Exp. 2) to compare the performance of the iAP-DNNs with six state-of-the-art patient-specific classifiers [4]–[9]. For fair comparisons, we followed the experimental protocols described in these studies. The purposes of these two experiments and how they use the MIT-BIH database are detailed as follows:

- **Exp. 1**: The first experiment aims to evaluate the performance of iAP-DNNs for classifying both Class S and Class V at the same time. To this end, we used 20 recordings (Record ID starting with Digit 1) for training the patient-independent DNN and another 24 recordings (Record ID starting with Digit 2) for adaptation and

testing. This means that we have 24 test patients and 24 patient-specific DNNs, each was adapted (either fine-tuning only or fine-tuning plus i-vector adaptation) by using the initial 5 minutes of his/her ECG recording. The remaining 25 minutes in the 24 recordings were used for performance evaluation.

- **Exp. 2**: The second experiment aims to evaluate the performance of iAP-DNNs in detecting S beats and V beats separately. To this end, we used 14 recordings (Record IDs 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233 and 234) for adaptation and testing of S-beat detection and 11 recordings (Record IDs 200, 202, 210, 213, 214, 219, 221, 228, 231, 233 and 234) for adaptation and testing of V-beat detection. As for the training, we used the remaining 30 recordings. Similar to Exp. 1, only the initial 5 minutes of these recordings were used for adaptation and the remaining were used for performance evaluation.

There are other studies [13]–[15] that use CNNs for ECG classification. We did not compare our results with the results in these studies for two reasons. First, [13]–[15] did not follow the AAMI standard, whereas all the methods to which we compared follow this standard. Second, [13] and [15] consider ECG classification as a binary classification problem, whereas we consider it as a multi-class classification problem.

Same as [4]–[9], the classification performance on each heartbeat class was measured by using four standard metrics, namely, classification accuracy (Acc), sensitivity (Sen), specificity (Spe) and positive predictive value (Ppv), which are calculated based on the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), as follows. Accuracy is the fraction of the total number of instances that is correctly identified, i.e., Acc = (TP+TN)/(TP+TN+FP+FN); sensitivity is the proportion of positives that are correctly identified, i.e., Sen = TP/(TP+FN); specificity is the proportion of negatives that are correctly identified, i.e., Spe = TN/(TN+FP); positive predictive value is the fraction of the positive predictions that are actually positive, i.e., Ppv = TP/(TP+FP). Details on how to interpret these four metrics can be found in [27]–[29].

Matthews correlation coefficient (MCC) [30], [31] was also calculated to measure the performance of different classifiers. MCC can reflect the performance of classifiers under severe data-imbalance scenarios. Receiver operating characteristics (ROCs) [32] were used to show the tradeoff between the performance measures (i.e., Sen vs. Spe) of a binary classifier when the decision threshold varies. Because the threshold typically has a wide range, ROC curves can provide more comprehensive information on performance.

C. DNN Structure and DNN Training

The general classifier has three hidden layers with a structure 417–100–100–100–5. The Glorot uniform initializer [33] was used to initialize the weights of the patient-independent DNN and the upper layers of the patient-specific DNNs. We used the rectified linear unit (ReLU) in the hidden layers. The Adam optimizer [34] with default parameters was used for stochastic mini-batch (batch size of 128) gradient descent. Batch normalization and dropout were employed to train the DNNs. A dropout layer was added between the input and the first hidden layer, and the dropout rate was set to 20%. In addition, 30% of the training set was reserved for validating the performance of the network after every epoch, so that early stopping can be applied to prevent overfitting. The early stopping strategy provides guidance on how many iterations should be run before the model begins to overfit the training data. The maximum number of epochs used for both patient-independent training and patient-specific training was set to 50. To train the i-vector extractor, we investigated different numbers of mixture components in the UBM (e.g., 16 and 20) and different i-vector dimensions (e.g., 32, 64 and 128), and the optimal combination was found to be 20 and 64 for the number of mixtures and i-vector dimension, respectively.

TABLE III: Performance of iAP-DNNs, with the i-vector being injected into different hidden layers of the network (Fig. 3(b)). "Correctly Classified" represents the number of correctly classified beats.

AAMI class		Bottom H	Middle H	Top H
N	Correctly classified	1178	1165	1109
	Ground truth	1193		
S	Correctly classified	25	41	37
	Ground truth	126		
V	Correctly classified	0	167	0
	Ground truth	198		
F	Correctly classified	0	0	0
	Ground truth	2		
Q	Correctly classified	0	0	0
	Ground truth	0		
Accuracy(%)		79.2	**90.4**	75.4

V. Performance Investigation

This section first investigates the best layer for injecting the i-vectors. Next, we demonstrate the effectiveness of i-vector adaptation. Finally, we compare the classification performance of iAP-DNNs with that of existing patient-specific classifiers.

A. Inject I-vector into Different Hidden Layers

Table III provides the classification accuracies of the iAP-DNNs. The results show that the performance was the best when the i-vector was injected into the middle hidden layer. Therefore, our justification in Section III-D is supported and this settings was applied to subsequent experiments.

B. Effect of I-vector Adaptation

To show the effect of i-vector adaptation, we created a patient-specific DNN by applying backpropagation fine-tuning on the patient-independent DNN (Fig. 3(a)) using 5 minutes of heartbeat vectors from a patient (e.g., Record ID 221). We also created a patient-specific iAP-DNN by applying backpropagation fine-tuning on the DNN in Fig. 3(b), not only using the 5 minutes of heartbeat vectors but also an i-vector extracted from the 5-minute heartbeats. Then, we presented ten minutes of ECG, including the normal and arrhythmic heartbeats of this patient, to both DNNs. Note that the five

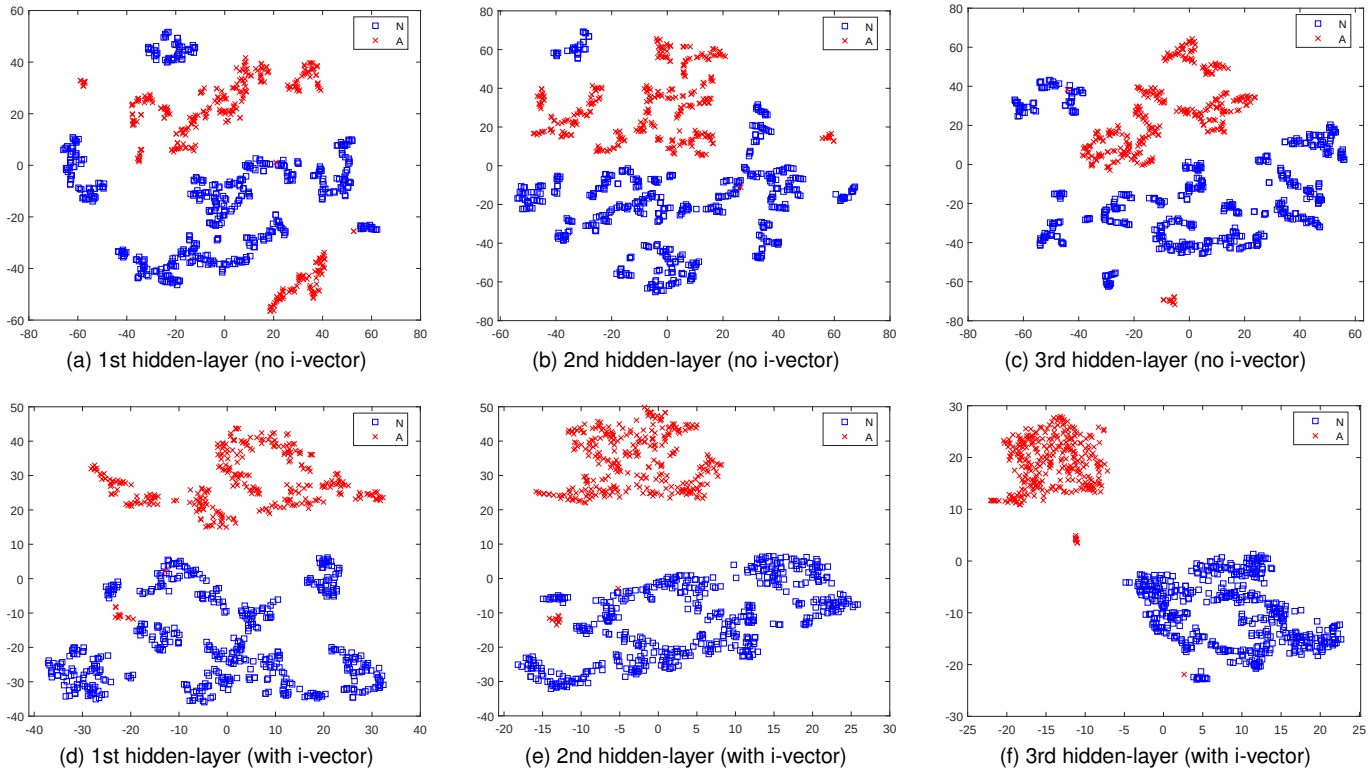


Fig. 6: t-SNE plots of the neuron activations at different hidden layers: (a)–(c) with patient's 5-minute ECG adaptation; (d)–(f) with patient's 5-minute ECG and i-vector adaptation. It is clear that with i-vector adaptation, the number of clusters is smaller and the A and N classes are well separated in (d)–(f).

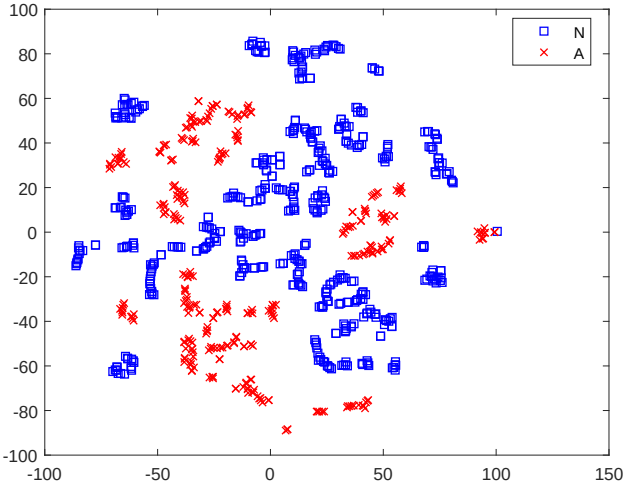


Fig. 5: t-SNE plot of 417-dimensional feature vectors. Squares (in blue) and crosses (in red) refer to normal heartbeats (N) and arrhythmias (A) of a patient, respectively.

minutes of ECG recordings comprise a majority of (but not necessarily all) ECG types of that patient. As different patients have different health conditions, the numbers of heartbeats for individual classes are also different.

The t-SNE plot of 417-dimensional feature vectors is shown in Fig. 5, where □ and × represent the normal (N) and arrhythmic (A) heartbeats, respectively. We can see that there is no obvious clusters in Fig. 5. We progressively moved up the hidden layers and projected the activations (before the ReLU) at the first, second and third hidden layers onto two-dimensional t-SNE spaces. The projected activations are shown in Fig. 6. Obviously, without i-vector adaptation Fig. 6(a)–(c), the projected vectors of both heartbeat types scatter in different regions of the t-SNE space and form multiple clusters, which makes classification more difficult. On the other hand, with i-vector adaptation (Fig. 6(d)–(f)), the two heartbeat types are well separated, which makes classification by the softmax layer easy. Moreover, from Fig. 6(d) to Fig. 6(f), we can see that each class has fewer clusters and the clusters of the two classes become more separate. This means that from the bottom to the top layers, the representation becomes more and more discriminative.

C. Classification Performance

1) Experiment 1 (Exp. 1): The first experiment was conducted to evaluate the proposed method based on 24 ECG recordings. Table IV shows the confusion matrix of iAP-DNNs. We can see that the performance is better if patients' i-vectors were used for adaptation. Specifically, the numbers of true positives for Class S and Class V have been increased. Besides, the performance of the iAP-DNNs and that of [4]–[6], [8], [9] are shown in Table V. Except for the Ppv of Class S and the Sen of Class V in [9], the overall performance of the proposed method for Class S and Class V is significantly better than that in [4]–[6], [8], [9] for all evaluation measures.

Using the confusion matrix in Table IV, the MCC performance of Classes N, S, V, F and Q can be calculated. Table

TABLE VII: Performance of the patient-specific classifiers in [4], [5], [7]–[9] and our iAP-DNNs (Exp. 2)

Method		[4]	[5]	[7] Method I	[7] Method II	[8]	[9]	iAP-DNNs
Class S	Acc	97.5	96.1	**99.1**	98.3	97.3	98.6	**99.1**
	Sen	74.9	81.8	76.5	61.4	**85.3**	77.2	78.4
	Spe	98.8	98.5	**99.9**	99.8	98.0	99.8	**99.9**
	Ppv	78.8	63.4	**99.1**	90.7	71.8	96.6	98.7
Class V	Acc	98.8	97.9	**99.7**	99.4	99.1	98.7	**99.7**
	Sen	94.3	90.3	97.1	91.8	96.4	97.2	**97.4**
	Spe	99.4	98.8	**99.9**	**99.9**	99.5	98.9	**99.9**
	Ppv	95.8	92.2	**98.5**	98.0	96.4	92.1	97.8

TABLE IV: Confusion matrix of iAP-DNNs in Exp. 1. The values in parentheses correspond to fine-tuning the DNN without i-vector injection.

	N	S	V	F	Q
N	41600 (41630)	78 (77)	92 (56)	47 (29)	4 (29)
S	439 (523)	1829 (1749)	63 (50)	4 (12)	2 (3)
V	225 (305)	69 (349)	4473 (4097)	39 (47)	1 (9)
F	64 (86)	2 (1)	49 (43)	496 (481)	0 (0)
Q	5 (5)	0 (1)	2 (1)	1 (1)	0 (0)

TABLE V: Performance of the patient-specific classifiers in [4]–[6], [8], [9] and the proposed iAP-DNNs (Exp. 1)

Method		[4]	[5]	[6]	[8]	[9]	iAP-DNNs
Class S	Acc	96.6	96.1	96.4	97.5	98.3	**98.7**
	Sen	50.6	62.1	64.6	76.8	68.7	**78.3**
	Spe	98.8	98.5	98.6	98.7	**99.8**	**99.8**
	Ppv	67.9	56.7	62.1	74.0	**94.7**	92.5
Class V	Acc	98.1	97.6	98.6	98.6	98.8	**98.9**
	Sen	86.6	83.4	95.0	93.8	**95.5**	93.1
	Spe	99.3	98.1	98.1	99.2	99.1	**99.5**
	Ppv	93.3	87.4	89.5	92.4	92.2	**95.6**

VI shows the performance comparison between the proposed iAP-DNNs and the existing patient-specific classifiers in [4]–[6], [8], [9]. Note that OMCC refers to overall MCC of the five classes. We can see that the MCC of the iAP-DNNs is much higher than the other three classifiers. The promising performance is not only found in the individual class, but also in the overall.

Fig. 7 shows the ROC curves of the proposed method for Class S and Class V. In the ROC curves, perfect classification (Spe = 1.0 and Sen = 1.0) corresponds to the upper right corner of the graph. A sensitivity-specificity operating point is good if it is close to the upper-right corner. In Fig. 7, the operating points of the best performing classifiers in [4]–[6], [8], [9] are also shown by the markers $+$, $\times$, $\circ$, $\square$, and $\bullet$, respectively. The figures clearly show that the sensitivity-specificity points in [4]–[6], [8], [9] are below the red curve. This means that, within a certain range of decision thresholds, the iAP-DNN achieves better performance in term of both sensitivity and specificity than the classifiers in [4]–[6], [8], [9].

2) Experiment 2 (Exp. 2): In the second experiment, for Class S and Class V, the evaluations were based on 14 and 11 test recordings, respectively. Table VII shows the Acc, Sen, Spe and Ppv of the iAP-DNNs and that of [4], [5], [7]–[9]. Note that in Method I of [7], five minutes of labeled ECG signals of a patient was used to adapt the patient-specific classifier. In Method II, the hypothesized labels were used instead of the manual labeling process. In Table VII, for Class V, the Sen of the iAP-DNNs is the highest among all methods and a high Spe (99.9%) is achieved. For Class S, although the Sen of the iAP-DNNs is lower than that in [8], its Spe and Ppv are higher.

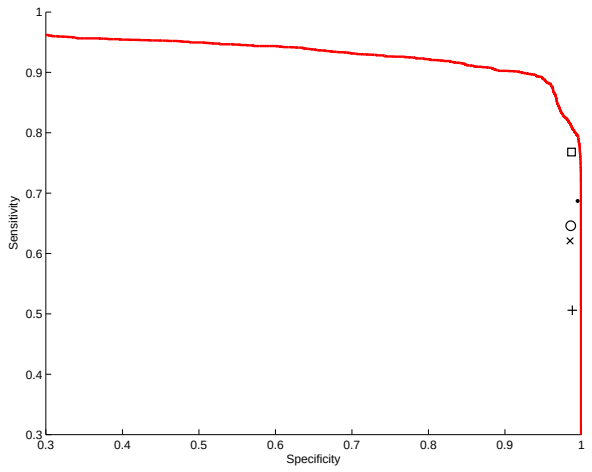
TABLE VI: Performance comparison in terms of MCCs (Exp. 1)

Method		[4]	[5]	[6]	[8]	[9]	iAP-DNNs
Class	N	0.83	0.81	0.84	0.88	0.90	**0.93**
	S	0.57	0.57	0.62	0.74	0.80	**0.84**
	V	0.87	0.83	0.91	0.92	0.93	**0.94**
	F	0.55	0.67	0.78	0.70	0.78	**0.83**
	Q	0.00	0.00	0.00	0.00	0.00	**0.00**
OMCC		0.93	0.92	0.94	0.95	0.96	**0.97**

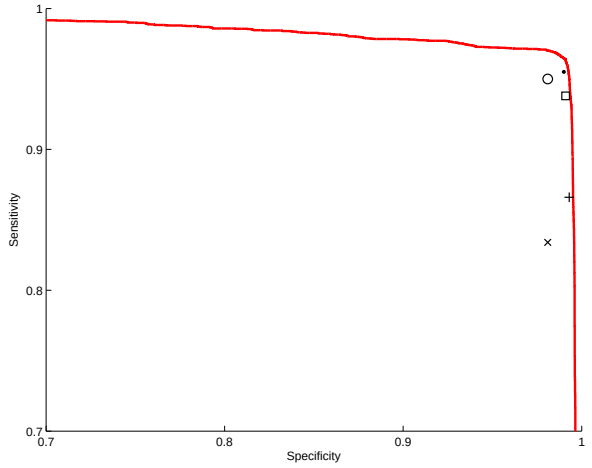
The performance of iAP-DNNs is similar to that of Method I in [7]. In [7], a subset was selected for training the general classifier based on the similarity among patients. The similarity is determined by calculating the dynamic time warping (DTW) distance, and the value of DTW threshold needs to be optimized by trial and error. However, in the proposed method, the ECG data of the general population can be used directly to train a general classifier before patient adaptation. Therefore, there is no need to throw away any ECG data from the general population nor do we need to optimize additional parameters. That is definitely an advantage.

VI. Conclusion and Future Work

In this paper, we introduce an adaptive patient-specific heartbeat classification model (i.e., iAP-DNNs) for diagnosing heart arrhythmias, which leverages the DNNs for both feature extraction and classification based on the raw ECG signals. A general classifier was first trained on the general population. Then, the weights in the lower part of the general classifier were retained and the weights in the upper part were randomized. To create a patient-specific classifier, and not only patient-specific ECG but also patient-dependent i-vectors are used for adaptation. Two experiments based on the MIT-BIH arrthymia database have been conducted. The results show that the proposed iAP-DNNs achieve better performance than existing patient-specific heartbeat classification systems.



(a) Class S $vs.$ non S (AUC = 0.946)



(b) Class V $vs.$ non V (AUC = 0.989)

Fig. 7: ROC curves (Sen $vs.$ Spe) of iAP-DNNs in Exp. 1. Black markers correspond to the best performance in [4]–[6], [8], [9]. *AUC*: Area under the ROC curve [35].

To the best of our knowledge, this is the first study that uses i-vectors to characterize the ECG of individual patients and applies the i-vectors to adapt a DNN for patient-specific ECG classification. The key contribution is that by injecting the i-vectors into a middle layer of the DNN during backpropagation fine-tuning, we can make the upper layers of the DNN more patient-dependent. Without the i-vectors as an auxiliary input to the middle layer, it is much harder to ensure such patient dependence.

A limitation of iAP-DNNs is that the method requires some patient-specific ECG data that have been manually labelled by medical doctors to adapt the patient-independent DNN. As long as the amount of adaptation data is small, this requirement will not pose a serious burden on the medical doctors nor the patients. However, for those patients without the access to medical services, the method is not applicable or they will need to fall back to using the patient-independent classifier. To relax such limitation, we can use an unsupervised adaptation approach as follows. For each patient, we use the patient-independent classifier to identify some highly discriminative heartbeats and hypothesize their labels. Then, we use these labels as the target outputs of the backpropagation algorithm to fine-tune the patient-independent classifier. As long as the patient-independent classifier is reliable enough, most of the hypothesized labels will be correct. This will be an interesting direction to pursuit in future work.

While the MIT-BIH arrhythmia dataset has been popular among the research community, it is also important to validate the accuracy using a larger dataset, e.g., the European ST-T Database [36]. This dataset consists of ninety two-hour ECG recordings with beats, rhythms, and signal quality annotation. We believe that the large amount of ECG data in this dataset is beneficial to the proposed method because it can leverage the data to train a better patient-independent classifier, which could lead to better patient-specific classifiers after i-vector adaptation.

REFERENCES

[1] Texas Heart Institute, "Holter Monitoring," https://www.texasheart.org/heart-health/heart-information-center/topics/holter-monitoring.
[2] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol.*, vol. 20, no. 3, pp. 45–50, May 2001.
[3] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Meth. Prog. Bio.*, vol. 127, pp. 144–164, Apr. 2016.
[4] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Trans. Neural Net.*, vol. 18, no. 6, pp. 1750–1761, Nov. 2007.
[5] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.
[6] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
[7] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "An automatic subject-adaptable heartbeat classifier based on multiview learning," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 6, pp. 1482–1492, Nov. 2016.
[8] X. Zhai and C. Tin, "Automated ECG classification using dual heartbeat coupling based on convolutional neural network," *IEEE Access*, vol. 6, pp. 27 465–27 472, June 2018.
[9] Y. Li, Y. Pang, J. Wang, and X. Li, "Patient-specific ECG classification by deeper CNN from generic to dedicated," *Neurocomputing*, vol. 314, no. 7, pp. 336–346, Nov. 2018.
[10] S. S. Xu, M. W. Mak, and C. C. Cheung, "Patient-specific heartbeat classification based on i-vector adapted deep neural networks," in *IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, Dec. 2018, pp. 784–787.
[11] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 4, pp. 788–798, Aug. 2011.
[12] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012.
[13] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Information Sciences*, vol. 415-416, pp. 190–198, Nov. 2017.
[14] Y. Hagiwara, H. Fujita, S. L. Oh, J. H. Tan, R. S. Tan, E. J. Ciaccio, and U. R. Acharya, "Computer-aided diagnosis of atrial fibrillation based on ECG signals: A review," *Information Sciences*, vol. 467, pp. 99–114, Oct. 2018.
[15] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, and R. S. Tan, "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals," *Applied Intelligence*, vol. 49, no. 1, pp. 16–27, Jan. 2019.
[16] Assoc. Adv. Med. Instrument., "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," *ANSI/AAMI*, no. EC57, 1998.
[17] S. W. Moon and S. G. Kong, "Block-based neural networks," *IEEE Trans. Neural Net.*, vol. 12, no. 2, pp. 307–317, Mar. 2001.

[18] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 308–311, May 2006.
[19] S. S. Xu, M. W. Mak, and C. C. Cheung, "Towards end-to-end ECG classification with raw signal extraction and deep neural networks," *IEEE J. Biomed. Health Inform.*, Sep. 2018. doi: 10.1109/JBHI.2018.2871510.
[20] M. W. Mak, "Lecture notes on factor analysis and i-vectors," Tech. Rep., Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University, 2016.
[21] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York Inc., 2006.
[22] L. V. D. Maaten and G. E. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[23] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[24] A. Mohamed, G.Hinton, and G. Penn, "Understanding how deep belief networks perform acoustic modelling," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012, pp. 4273–4276.
[25] T. N. Sainath, R. J. Weiss, A. Senior, K. W. Wilson, and O. Vinyals, "Learning the speech front-end with raw waveform CLDNNs," in *Proc. INTERSPEECH*, Sep. 2015, pp. 1–5.
[26] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–236, Mar. 1985.
[27] E. J. Boyko, "Ruling out or ruling in disease with the most sensitiue or specific diagnostic test: Short cut or wrong turn?" *Med. Decis. Making*, vol. 14, no. 2, pp. 175–179, Apr. 1994.
[28] A. Laupacis and N. Sekar, "Clinical prediction rules: a review and suggested modifications of methodological standards," *JAMA*, vol. 277, no. 6, pp. 488–494, Feb. 1997.
[29] A. K. Akobeng, "Understanding diagnostic tests 1: sensitivity, specificity and predictive values," *Acta Paediatr.*, vol. 96, no. 3, pp. 338–341, Mar. 2007.
[30] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochim. Biophys. Acta*, vol. 405, no. 2, pp. 442–451, Oct. 1985.
[31] M. W. Mak, J. Guo, and S. Y. Kung, "PairProSVM: protein subcellular localization based on local pairwise profile alignment and SVM," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 5, no. 3, pp. 416–422, Sep. 2008.
[32] J. A. Swets, *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.
[33] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *13th Int. Conf. on Artificial Intelligence and Statistics*, May 2010, pp. 249–256.
[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv*:1412.6980, Dec. 2014.
[35] T. Fawcett, "An introduction to roc analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
[36] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. 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, pp. 1164–1172, Apr. 1992.