The following publication S. S. Xu, M. -W. Mak and C. -C. Cheung, "Towards End-to-End ECG Classification With Raw Signal Extraction and Deep Neural Networks," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1574-1584, July 2019 is available at https://doi.org/10.1109/JBHI.2018.2871510.

Towards End-to-End ECG Classification with Raw Signal Extraction and Deep Neural Networks

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

***Abstract*—This paper proposes deep learning methods with signal alignment that facilitate the end-to-end classification of raw electrocardiogram (ECG) signals into heartbeat types, i.e., normal beat or different types of arrhythmias. Time-domain sample points are extracted from raw ECG signals, and consecutive vectors are extracted from a sliding time-window covering these sample points. Each of these vectors comprises the consecutive sample points of a complete heartbeat cycle, which includes not only the QRS complex but also the P and T waves. Unlike existing heartbeat classification methods in which medical doctors extract handcrafted features from raw ECG signals, the proposed end-to-end method leverages a deep neural network (DNN) for both feature extraction and classification based on aligned heartbeats. This strategy not only obviates the need to handcraft the features but also produces optimized ECG representation for heartbeat classification. Evaluations on the MIT-BIH arrhythmia database show that at the same specificity, the proposed patient-independent classifier can detect supraventricular- and ventricular-ectopic beats at a sensitivity that is at least 10% higher than current state-of-the-art methods. More importantly, there is a wide range of operating points in which both the sensitivity and specificity of the proposed classifier are higher than those achieved by state-of-the-art classifiers. The proposed classifier can also perform comparable to patient-specific classifiers, but at the same time enjoys the advantage of patient independency.**

***Index Terms*—ECG classification; arrhythmia classification; end-to-end; deep neural networks, heartbeat alignment.**

I. Introduction

Heart arrhythmias refer to the condition in which a patient's heart beats irregularly. Most types of arrhythmias have no symptoms and are not serious. However, arrhythmias may cause symptoms of heart diseases, including lightheadedness, passing out, shortness of breath and chest pain. Some types of arrhythmias such as atrial fibrillation, ventricular escape and ventricular fibrillation may cause strokes and cardiac arrest that are extremely dangerous and require immediate treatment [1].

Heart arrhythmias can be detected by using Electrocardiography (ECG), which records the electrical activities of a patient's heart using two electrodes attached to the skin. In general, an ECG recording session lasts several minutes, and medical doctors examine the ECG waveforms beat-by-beat to diagnose whether heart arrhythmias exist or not. The process is very tedious and time-consuming. Therefore, automatic heartbeat classification from ECG signals is important for diagnosing heart arrhythmias in medical practice.

A standard ECG refers to a 12-lead ECG and some ECG-based biometrics research [2], [3] was based on this conventional 12-lead configuration. During measurement, patients are asked to lie quietly on a bed so that high quality 12-lead ECG signals can be recorded, but this arrangement is impractical for long-term monitoring. A 2-lead configuration is routinely used in Holter monitoring [4] and is widely accepted as a practical means of long-term continuous heart monitoring. In this work, the proposed end-to-end heartbeat classification system is designed to detect some types of arrhythmias during long-term continuous heart monitoring. Thus, we used the 2-lead ECG configuration and also used MIT-BIH arrhythmia database [5] for performance evaluation because it comprises a standard set of Holter recordings for evaluating arrhythmia detectors.

In the last two decades, much research effort [6]–[17] has been spent on classifying heartbeats automatically. Many of these studies used the MIT-BIH arrhythmia database for performance evaluation. Most of these approaches involved three steps: preprocessing, feature extraction and classification. Preprocessing segments the heartbeats from the continuous ECG signals into individual beats.[1] Feature extraction converts the variable-length time-domain heartbeats into fixed-length feature vectors that encode the heartbeat's characteristics. Various features have been extracted from ECG signals to represent the heartbeats, such as Hermite coefficients [8], morphological features [7], wavelet transform features [9], [10], heartbeat interval features [7], [11] and sparse decomposition [17]. For classification, different machine learning algorithms have been investigated, including support vector machines (SVMs) [8], [10], [17], artificial neural networks (ANNs) [9], deep neural networks (DNNs) [11], convolutional neural networks (CNNs) [12], [14]–[16], multi-view-based learning [13], mixture-of-experts (MOEs) [6], and linear discriminants (LDs) [7]. Note that the performance of previous approaches cannot be compared directly because they used different standards to preprocess the ECG data. Only the works in [6], [7], [9], [10], [12], [13], [17] followed a well-known standard, i.e., American National Standard prepared by the Association for the Advancement of Medical Instrumentation (ANSI/AAMI EC57:1998) [18].

Hu *et al*. [6] proposed using MOEs for ECG classification. In this approach, each heartbeat comprises 14 sample points

S. S. Xu, M. W. Mak (enmwmak@polyu.edu.hk) and C. C. Cheung are with Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong SAR of China. This project was in part supported by the RGC of Hong Kong, Grant No. PolyU 1521371/17E.

[1] In this work, we focused on the beat-by-beat classification of ECG signals because most state-of-the-art ECG analysis algorithms [6]–[13], [16], [17] adopt the beat-by-beat analysis strategy.

on either side of its R peak. First, a classifier called the global classifier was trained based on a number of ECG recordings from different patients. Next, five minutes of doctor-annotated ECG signals obtained from a new patient were used for training a local classifier. Then, the global and local classifiers were combined to form a patient-specific MOE classifier. Although high accuracies have been achieved, the classifier cannot be generalized to all patients.

Chazal *et al*. [7] utilized morphological and dynamic features to represent heartbeats and then classified them into five classes. The classifier is based on linear discriminants, and its parameters are determined by maximum-likelihood estimation. In [10], Ye *et al*. applied wavelet transform and independent component analysis (ICA) to extract morphological features from segmented heartbeats. Heartbeat intervals were also used as dynamic features. The features were applied to an SVM for classifying heartbeats into five classes. While the classification processes in [7] and [10] are completely automatic and patient-independent, the accuracies of their classifiers are significantly lower than that of [6]. This is due to variability in ECG characteristics among different patients.

In [9], Ince *et al*. applied wavelet transform and principal component analysis (PCA) to extract morphological features, which were then combined with temporal features to form the final feature vectors. A multidimensional particle swarm optimization method was proposed to train an artificial neural network (ANN) based classifier using 245 common training beats and a variable number of patient-specific beats.

In [12], the raw data of each beat in the MIT-BIH arrhythmia database were downsampled to 64 or 128 time-points centered on the R-peak. The time-domain signals or their FFT representations are used as the input to patient-specific 1-D CNN. Each CNN was trained by 245 representative beats that are common to all patients and 5 minutes of patient-specific beats that are specific to the corresponding patient. The CNNs were evaluated by using the remaining 25 minutes of the patient's ECG recording. Results showed that except for Class S, the CNNs outperform any existing arrhythmia classifiers (including their earlier work in [9]) on the same dataset. The high performance, however, relies on the manual annotations of the 5-minute patient-specific data by medical doctors. Furthermore, the CNN classifiers are patient-dependent, meaning that any new patients wanting to use this technology need to provide 5 minutes of ECG recordings, followed by an expensive annotation process.

Ye *et al*. [13] proposed a subject-adaptable heartbeat classification model to overcome the problem of interpersonal variations in ECG signals. The preprocessing and feature extraction methods in [13] are identical to those in their earlier work [10]. However, unlike [10], the classification models in [13] are customized for each patient; also, unlike other patient-specific classifiers, the subject-customized classifiers in [13] can be trained on unlabeled patient-specific data, meaning that no manual intervention is required during training. This is achieved by dividing the classification models into a general classifier and a specific classifier for each patient. The former is trained on the data extracted from patients who are similar to the target patient, whereas the latter is trained on a small set of patient-specific heartbeats with high-confidence labels hypothesized by multi-view learning models (multi-channels plus temporal information). During classification, the decisions of the general and specific classifiers are probabilistically combined so that both the inter-patient and intra-patient perspectives of the classification task are considered in the final decisions.

In [17], a new feature extraction method (sparse decomposition over a Gabor dictionary) is proposed to represent various classes of heartbeats. Four kinds of features (i.e., time delay, frequency, width parameter and square of expansion coefficient) are extracted from each of the significant atoms of the dictionary and concatenated to constitute a feature vector. The feature vectors are classified using some typical classification models. Among the different proposed methods, the performance of the particle swarm optimization (PSO) optimized least-square twin SVM model achieves the best performance.

In this paper, we evaluated the performance of the proposed method on the MIT-BIH arrhythmia database and followed the ANSI/AAMI EC57 standard to compare with those in [7], [9]–[13], [17]. Study [11] was chosen for performance comparison because it uses DNNs as classifiers. We did not compare our results with [6] because its performance is poorer than that of [9], [12], [13]. A comparative summary of these studies [6]–[17] as well as our proposed method is shown in Table I.

We propose an end-to-end method with a deep neural network (DNN) for both feature extraction and classification based on aligned heartbeats. This method obviates the need to handcraft the features and produces optimized ECG representation for heartbeat classification. Through the performance investigation using the MIT-BIH arrhythmia database, the proposed method performs better than current state-of-the-art methods. The paper is organized as follows. Section II introduces the proposed end-to-end ECG classification algorithm. Section III outlines the experimental setting. Section IV compares the performance of the end-to-end ECG classifier against existing ECG classification methods. Finally, Section V concludes our findings.

II. End-to-End ECG Classification

In this section, we first explain the motivation to build an end-to-end ECG classifier, and then provide a system overview of the classifier. Next, we describe the deep neural network inside the classifier, and finally, explain the heartbeat segmentation and alignment procedures in the classifier.

A. Motivation

In most previous works [6]–[13], handcrafted feature[2] vectors were extracted from the QRS complex of heartbeats because this region is thought to contain most ECG pulse informantion. However, studies [20]–[22] show that the P and T waves also contain important information relevant to heart arrhythmias. In light of this observation, we proposed to use

[2]The term "handcrafted features" is frequently used in the machine learning community to refer to features that are handcrafted by human experts of the field based on their knowledge and past experience.

TABLE I: Summary of the published studies [6]–[17] and the proposed method. (a) Studies that do not follow the ANSI/AAMI EC57 standard. (b) Studies that follow this standard. All studies in (a) use the class-oriented evaluation scheme. Therefore, their classifiers are neither patient-specific nor patient-independent.

Ref.	Classes	Database	Analysis Strategy	Evaluation Scheme	Features	Classifier	Classifier Tpye
Osowski [8]	13	MITDB	Beat-by-beat	Class-oriented	Hermite	SVM	N/A
Jun [11]	2	MITDB	Beat-by-beat	Class-oriented	Morphological, RR-interval	DNN	N/A
Acharya [14]	4	MITDB, CUDB [19], AFDB [19]	Segments	Class-oriented	Downsampling, Wavelet, Sample points	CNN	N/A
Tan [15]	2	PhysioNet [19]	Segments	Class-oriented	Upsampling, Wavelet, Sample points	CNN-LSTM	N/A
Acharya [16]	2	PTB [19]	Segments	Class-oriented	Raw ECG	CNN	N/A

(a)

Ref.	Classes	Database	Analysis Strategy	Evaluation Scheme	Features	Classifier	Classifier Tpye
Hu [6]	4	MITDB	Beat-by-beat	Subject-oriented	Downsampling, Sample points	MOE	Patient-specific with expert intervention
Chazal [7]	5	MITDB	Beat-by-beat	Subject-oriented	Morphological, RR-interval	LD	Patient-independent
Ince [9]	5	MITDB	Beat-by-beat	Subject -oriented	Wavelet, PCA, RR-interval	ANN	Patient-specific with expert intervention
Ye [10]	5	MITDB	Beat-by-beat	Subject-oriented	Wavelet, ICA, RR-interval	SVM	Patient-independent
Kiranyaz [12]	5	MITDB	Beat-by-beat	Subject-oriented	Downsampling, FFT	CNN	Patient-specific with expert intervention
Ye [13]	5	MITDB	Beat-by-beat	Subject-oriented	Wavelet, ICA, RR-interval	Multi-view Learning	Patient-specific without expert intervention
Raj [17]	5	MITDB	Beat-by-beat	Subject-oriented	Sparse	Least-square Twin SVM	Patient-independent
Proposed method	**5**	**MITDB**	**Beat-by-beat**	**Subject-oriented**	**Raw ECG**	**DNN**	**Patient-independent**

(b)

raw ECG waveforms as the input of a deep neural network (DNN) classifier, which we refer to as end-to-end ECG classification. The advantage of using raw ECG waveforms is that the QRS complex and P and T waves can be included in the extracted heartbeats so that better representations can be obtained for classification.

CNN is another network type that can receive raw ECG signals as input. One advantage of CNNs is that they are invariant to time-domain translation. Even with this translation-invariant property, some CNN-based methods (e.g., [12], [16]) still require detecting the R peaks to fulfill the requirements of beat-by-beat analysis. Only the methods in [14], [15] do not require R peak detection because they segment the ECG signals into fixed time intervals of 2 to 5 seconds. The fixed time intervals may contain one or several heartbeat cycles, and ECG segments are not applicable to the beat-by-beat analysis strategy. Furthermore, the results in [14]–[16] were based on others' database and did not follow the ANSI/AAMI EC57 standard. Therefore, we did not compare the proposed method with [14]–[16] in this work.

While both of our proposed DNNs and the CNNs in [12], [14]–[16] use raw ECG signals as input, our raw signal extraction method has two advantages over them. First, instead of simply cropping equal numbers of time points from the left and right of an R-peak as in [16], we align the heartbeats to ensure that the input to the DNN contains the QRS complex, the P wave and the T wave. Second, to fix the input dimension, the method in [12], [14], [15] upsamples or downsamples the raw ECG signals to certain time-points per beat, which may cause information loss. In contrast, our alignment method allows the DNN to fully utilize the information in the ECG signals by keeping more time-points (417 in this paper, which will be described in Section II-D) per beat.

B. System Overview

This paper proposes an end-to-end ECG classification system shown in Fig. 1. The system receives raw ECG signals at one end and produces beat-by-beat classification decisions at the other end. In the figure, preprocessing refers to the process of extracting heartbeats from continuous ECG signals, which involves heartbeat segmentation and alignment. The DNN in Fig. 1 is used for both feature extraction and classification, which are achieved by the lower part and the upper part of the network, respectively. The design of the DNN is discussed in the Section II-C.

To extract fixed-length feature vectors from raw ECG signals, two steps must be performed: (1) heartbeat segmentation and (2) heartbeat alignment. These two steps will be described in the Section II-D and II-E respectively.

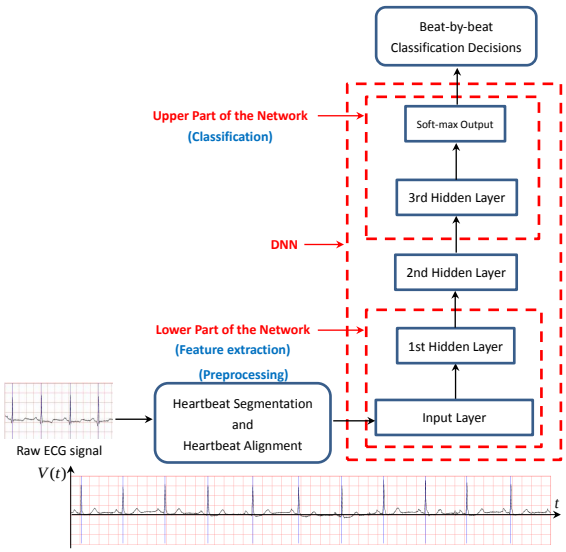


Fig. 1: End-to-End heartbeat classification system.

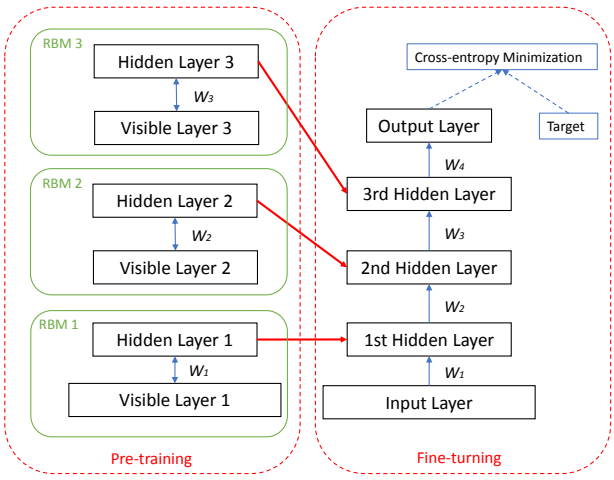


Fig. 2: DNN with stacked RBMs.

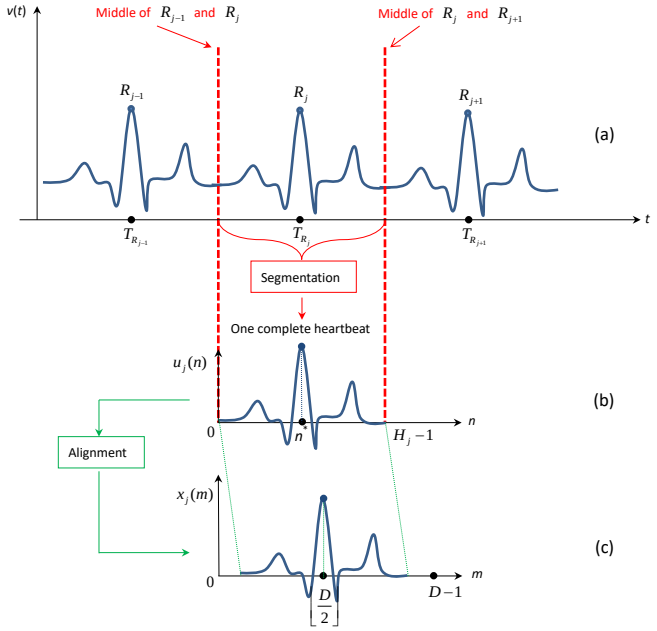


Fig. 3: Hypothetical example illustrating the heartbeat segmentation and alignment processes. In (c), $\lfloor a \rfloor$ means the integer (floor) of a.

C. Design of Deep Neural Networks

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

$$y_k = \frac{\exp\left\{a_k^{(L)}\right\}}{\sum_{j=1}^{K} \exp\left\{a_j^{(L)}\right\}}, \qquad k = 1, \ldots, K. \tag{1}$$

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

The weights in the hidden layers can be pre-trained by a greedy layer-wise unsupervised training process [23] in which each hidden layer is considered as a restricted Boltzmann machine (RBM) [24], [25] whose weights are optimized by the contrastive divergence algorithm [26]. Alternatively, the weights can be initialized by the Xavier initializer [27]. Then, the backpropagation algorithm is used to fine-tune the whole network by minimizing the cross-entropy error between the target outputs and the actual outputs:

$$E_{\text{ce}} = -\sum_{n}\sum_{k=1}^{K} t_{n,k} \log y_{n,k}, \tag{2}$$

where $y_{n,k}$ is the actual output of node k, n indexes the training vectors in a mini-batch, and $t_{n,k} \in \{0, 1\}$ are the target outputs which follow the one-hot encoding scheme.

In this work, we used a DNN with stacked RBMs as shown in Fig. 2. The RBM at the bottom layer has Gaussian visible nodes and Bernoulli hidden nodes. The remaining RBMs have Bernoulli distributions in both visible and hidden layers. During fine-tuning, the pre-trained weights (W_1, W_2 and W_3) were used as the initial weights and the weights between the upper two layers (W_4) were initialized with small random numbers. In addition, 30% of the training set was used for computing the accuracy of the network after every epoch, so that early stopping can be applied to prevent overfitting. Note that the pre-training step can provide necessary regularization to the network [28] and the early stopping strategy provides guidance on how many iterations should be run before the model begins to over-fit the training data.

D. Heartbeat Segmentation

The bottom of Fig. 1 shows a continuous ECG signal in the MIT-BIH arrhythmia database. To extract a complete heartbeat from the ECG signal, we need to define what a complete heartbeat is and then perform heartbeat segmentation. Since the R peak usually occurs around the middle of a heartbeat, we can use it as an anchor point for locating a complete heartbeat.

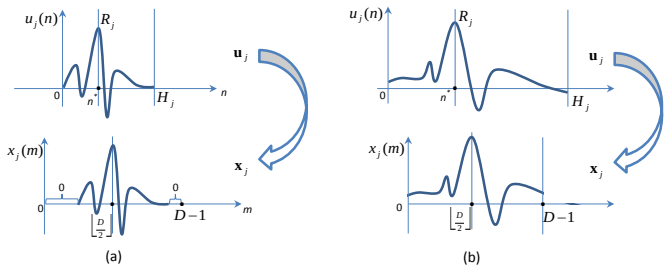


Fig. 4: Creating feature vector $\mathbf{x}_j$ from $\mathbf{u}_j$ by aligning sample $u_j(n^*)$ to the midpoint of $\mathbf{x}_j$. (a) Example of zero-padding, $H_j < D$. (b) Example of truncation, $H_j > D$.

The positions of R peaks can be accurately determined (over 99%) by using Pan-Tompkins algorithm in [29]. We assume that the R peak is located at the center of its corresponding heartbeat, and thus the boundary of a complete heartbeat is assumed to lie on the middle of two successive R peaks. Based on this assumption, a complete heartbeat comprises the sample points between the two middle points of three consecutive R peaks. Fig. 3(a) shows an example of a complete heartbeat and its relationship with its preceding and succeeding heartbeats. In Fig. 3(a), t indexes the sample points of an ECG signal, $v(t)$ is the voltage (in mV) of the ECG signal at time index t, R_j is the j-th R peak, and T_{R_j} is the time index of R_j.

After heartbeat segmentation, we obtain the j-th complete heartbeat $\mathcal{H}_j$, which is an integer set containing sample points between $\lfloor \frac{1}{2}(T_{R_{j-1}} + T_{R_j}) \rfloor$ and $\lfloor \frac{1}{2}(T_{R_j} + T_{R_{j+1}}) \rfloor$, where $\lfloor a \rfloor$ means the integer (floor) of a. As illustrated in Fig. 3(b), the elements in $\mathcal{H}_j$ are indexed by $n = 0, \ldots, H_j - 1$, where H_j is the number of sample points in the complete heartbeat. More precisely, we have

$$H_j = \left\lfloor \frac{1}{2}(T_{R_j} + T_{R_{j+1}}) \right\rfloor - \left\lfloor \frac{1}{2}(T_{R_{j-1}} + T_{R_j}) \right\rfloor + 1. \quad (3)$$

We may use a vector $\mathbf{u}_j$ to represent $\mathcal{H}_j$ as follows:

$$\mathbf{u}_j = [u_j(0), \ldots, u_j(n^*), \ldots, u_j(H_j - 1)]^\mathsf{T}, \quad (4)$$

where $n^* = T_{R_j} - \lfloor \frac{1}{2}(T_{R_{j-1}} + T_{R_j}) \rfloor$ is the time index corresponding to the peak in $\mathbf{u}_j$.

However, $\mathbf{u}_j$ still cannot be directly used for training a DNN because the number of sample points is not a constant (the duration of each complete heartbeat is not the same). A fixed number of samples (D) needs to be set for each heartbeat. Thus, we measured the durations of all segmented heartbeats and found a value that is larger than 95% of all durations. In our experiments, D was found to be 417 and this value was applied to all of the completed heartbeats.

E. Heartbeat Alignment

Because we use the R peak as the anchor point of a heartbeat in the heartbeat segmentation process, it is necessary to align it to the midpoint of the D consecutive time points of each heartbeat. Fig. 3(b) and Fig. 3(c) show the alignment process. We extract samples from $\mathbf{u}_j$ in Eq. 4 to produce a feature vector

$$\mathbf{x}_j = [x_j(0), \ldots, x_j(D-1)]^\mathsf{T} \quad (5)$$

such that the $\lfloor \frac{D}{2} \rfloor$-th element in $\mathbf{x}_j$ is aligned to n^*-th element in Eq. 4. Note that this procedure requires zero padding and sample truncation for most heartbeats. Specifically, when $H_j > D$, we may need to truncate some of the samples in the head or tail or both the head and tail of $\mathbf{u}_j$. However, when $H_j < D$, we may need to pad zeros to the head or tail or both the head and tail of $\mathbf{u}_j$. In some rare cases, both zero padding and sample truncation need to be performed. Fig. 4 shows some examples of the alignment process. Given Eq. 4 and Eq. 5, the alignment process can be implemented as follows:

$$x_j(m) = \begin{cases} 0, & \text{if} \quad m < \lfloor \frac{D}{2} \rfloor - n^* \\ & \text{or} \quad m > \lfloor \frac{D}{2} \rfloor + (H_j - n^*) \\ u_j(m - \lfloor \frac{D}{2} \rfloor + n^*), & \text{otherwise} \end{cases} \quad (6)$$

where $m = 0, 1, \ldots, D-1$.

After heartbeat segmentation and alignment, the set of feature vectors in a dataset is denoted as

$$\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_N\}, \quad (7)$$

where $\mathbf{x}_1$ and $\mathbf{x}_N$ correspond to the second and the second last beats in a record, and N is the number of complete heartbeats.

The process of heartbeat alignment is vital to the high performance of the end-to-end DNN (see results in Section IV). Because the DNN receives time-domain ECG signals as input, its internal structure represents not only the pulse shapes of heartbeats but also their relative positions along the time axis. Without the R-peak alignment, the R peak in Fig. 3(c) could be in many possible locations, causing high variability in the feature vectors. By aligning the R peak to the mid-point of $x_j(m)$ in Fig. 3(c), we essentially make the DNN invariant to the phase shift of the ECG signals.

III. Experimental Setting

In this section, we first briefly describe the data set, and then introduce our evaluation scheme and some issues concerning its implementation (i.e., evaluation protocol). Finally, since DNN performance is greatly affected by its network structure, we describe how we find the optimized network structure of the DNN used in the classifier.

A. Data Set

The MIT-BIH arrhythmia database [5] contains 48 half-hour excerpts of 2-lead ambulatory ECG recordings. It involves 47 subjects: 25 men aged between 32 and 89, and 22 women aged between 23 and 89. Each record contains a continuous recording of ECG signals from a single subject, except for Records 201 and 202 in which the data were obtained from the same male subject. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10-mV range. All records contain raw ECG signals, and they were independently labelled beat-by-beat by two or more cardiologists. The total number of labelled heartbeats is 108,655. These heartbeats are divided into 15 different types.

Table II shows the 15 types of heart arrhythmia in the MIT-BIH arrhythmia database. According to the American National Standard (ANSI/AAMI EC57:1998) prepared by the

TABLE II: Heartbeat types in the MIT-BIH arrhythmia database [5]

Class Code	Full Name	No. of Instances
NOR	Normal beat	74,478
LBBB	Left bundle branch block beat	8,074
RBBB	Right bundle branch block beat	7,259
AP	Atrial premature beat	2,544
aAP	Aberrated atrial premature beat	150
NP	Nodal (junctional) premature beat	83
SP	Supraventricular premature beat	2
PVC	Premature ventricular contraction	6,901
fVN	Fusion of ventricular and normal beat	802
AE	Atrial escape beat	16
NE	Nodal (junctional) escape beat	215
VE	Ventricular escape beat	106
P	Paced beat	7,028
fPN	Fusion of paced and normal beat	982
U	Unclassifiable beat	15

TABLE III: Mapping the MIT-BIH arrhythmia types into five heartbeat classes recommended by AAMI

	AAMI Class				
	N	S	V	F	Q
MIT-BIH Class Code (see Table II)	NOR, LBBB, RBBB, AE, NE	AP, aAP, NP, SP	PVC, VE	fVN	P, fPN, U
No. of Instances	90,042	2,779	7,007	802	15

Association for the Advancement of Medical Instrumentation, these heartbeat types can be combined into five classes as shown in Table III. These classes include normal beat (N), ventricular ectopic beat (V), supraventricular ectopic beat (S), fusion of a normal and a ventricular ectopic beat (F) and unknown beat type (Q).

In most of the recordings in the MIT-BIH database, the upper signal is a modified limb lead II (MLII), and the lower one is a modified lead V1.[3] In our experiments, only the upper signal was used for ECG classification because normal QRS complexes are usually prominent in it.

B. Evaluation Scheme

As mentioned in [10], two evaluation schemes, namely "class-oriented" and "subject-oriented", are commonly used for ECG classification. Using the class-oriented evaluation scheme, the performance of the classifiers in [8], [11], [14]–[16] may be overestimated because signals in the training and test sets could belong to the same patient. The "well trained" classifier may fail to predict the ECG signals from an unseen individual. The scheme is not applicable in practice because of the significant variation in ECG characteristics among different subjects. Using the subject-oriented evaluation scheme, the data in [6], [7], [9], [10], [12], [13], [17] were divided into the training and the testing set based on ECG recordings. This means that the ECG signals in the training and test set were definitely not from the same patient. The classifiers produced by this scheme perform more realistically.

The subject-oriented evaluation scheme leads to two types of classifiers—patient-independent classifiers (e.g., [7], [10], [17]) and patient-specific classifiers (e.g., [6], [9], [12], [13]). In general, patient-specific classifiers perform much better than patient-independent classifiers because the formers are trained on a small set of annotated data from the respective patients. In contrast, the cost of patient-independent classifiers is much lower because no patient-specific data or expert intervention is required. Note that the proposed end-to-end method adopts the subject-oriented evaluation scheme, and a patient-independent classifier is built for beat-by-beat classification of ECG signals.

C. Evaluation Protocol

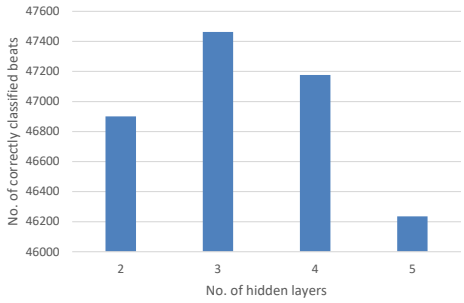
In compliance with the AAMI recommended practice, four recordings containing paced beats were removed from the dataset. The remaining 44 records were split into two datasets (DS1 and DS2),[4] with each dataset containing approximately 50,000 beats from 22 recordings. Note that this way of splitting the data had also been used in [7], [10] and [17]. Following their evaluation protocols (the subject-oriented evaluation scheme), we applied 22-fold cross validation on DS1 in one experiment (Exp. 1) and used DS1 as the training set and DS2 as the test set in another experiment (Exp. 2). Note that in Exp. 1, each record was used as test data in sequence and the other 21 records were used as training data. Such process was repeated 22 times so that each record had been used once as the test data. As a result, we may compare our results with previous studies.

As mentioned before, there are two evaluation schemes and this paper follows the "subject-oriented" scheme because it is more realistic. All research works (in Table I(b)) following this scheme use ANSI/AAMI EC57 standard with MITDB for performance investigation. As suggested by the ANSI/AAMI EC57 standard [18], we focused on evaluating the classification performance of the two majority arrhythmia classes (Classes S and V). Among the performance indicators for medical diagnoses, sensitivity (SEN) and specificity (SPC) are two important measures of the diagnostic accuracy of a test because a highly sensitive test can be useful for ruling out a disease if a person has a negative result, whereas a highly specific test can be useful for ruling in patients who have a certain disease. Some medical publications [30], [31] recommend clinicians to choose the most sensitive diagnostic test to rule out disease and the most specific diagnostic test to rule in disease. Therefore, in this work, the diagnostic performance on Class S and Class V was measured using SEN and SPC. Besides, because the overall accuracy measures the overall system performance over all classes, it was also used in this work.

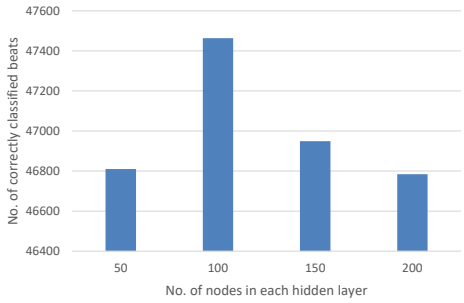
We also used the receiver operating characteristics (ROCs) [32] to show the performance on Class S and Class V of

[3]We adopted the terminology from MIT-BIH and used upper and lower signals to refer to the two channels of ECG recordings.

[4]DS1 contains data from ECG recordings 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 and 230; DS2 contains data from ECG recordings 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233 and 234.



(a) Increase hidden layers



(b) Different number of hidden nodes

Fig. 5: Network structure optimization

the end-to-end classifier. An ROC curve shows the tradeoff between two performance measures (e.g., sensitivity versus specificity) 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.

In addition, Table III shows the number of classes is highly imbalanced. Matthews correlation coefficient (MCC) [33], [34] is a better measure for imbalanced datasets. MCC is a value between -1 and $+1$. A coefficient of "$+1$" represents a perfect prediction, a "0" means it is not better than random prediction, and a "-1" indicates total disagreement between prediction and the ground truth.

D. Network Structure

Fig. 5(a) shows the effect of increasing the number of hidden layers on the network's ability to classify heartbeats. 22-fold cross validation was applied to DS1 and the total number of heartbeats used for evaluation was 50,977. The results show that the performance was the best when the number of hidden layer was 3. To further optimize the network structure, we fixed the number of hidden layers to 3 and varied the numbers of hidden nodes (e.g., 50, 100, 150 and 200) per layer. According to Fig. 5(b), the best performance was obtained when the number of hidden nodes equaled to 100. Therefore, in subsequent experiments, the DNN classifier contained 3 hidden layers and each layer had 100 nodes.

IV. Performance Investigation

This section first presents the feature extraction capability of our proposed method, then it describes and compares our method's performance with current start-of-the-art methods. Next, we compare our proposed method with other DNN classifiers including comparing our proposed method with patient-specific classifiers. Finally, we summarize our findings in this section.

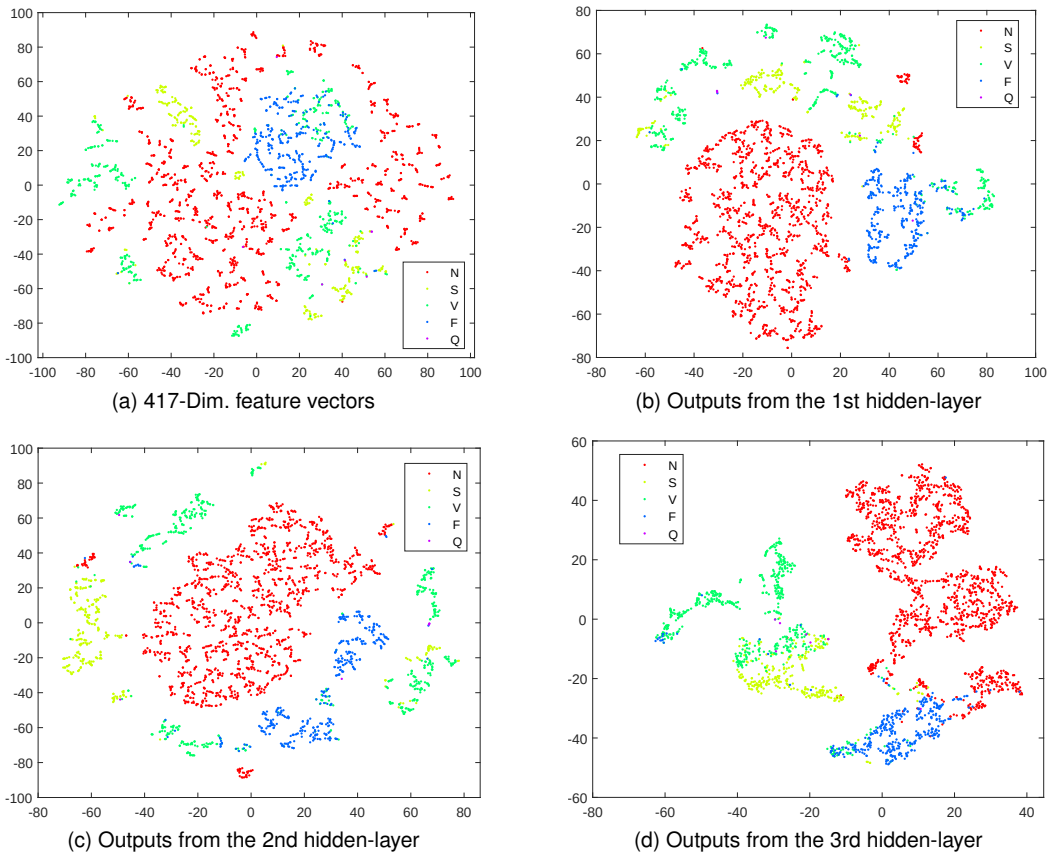
A. Hidden Node Representation

The t-distributed stochastic neighbor embedding (t-SNE) [35] is a nonlinear dimension reduction method for visualizing high-dimensional data on a two or three-dimensional space. Using the 417-dimensional vectors as input, we extracted the outputs from the first, second and third hidden layers of the DNN. We applied t-SNE on the 417-dimensional vectors and different hidden layers. The results are shown in Fig. 6 for Classes N, S, V, F and Q. To allow a good visualization, the number of samples of Class N is reduced in the figures. No obvious clusters can be observed in the feature vectors (Fig. 6(a)). When we progressively move up the hidden layers (Figs. 6(b)–(c)), the clustering property becomes apparent. However, in the first two hidden layers, each class still has multiple clusters, meaning that further nonlinear operations are required. In the third hidden layer (Fig. 6(d)), the clusters are very obvious, and more importantly each class has fewer clusters and the clusters of different classes become more separated. This means that from the bottom to the top layers, the representation becomes more and more discriminative. From another perspective, the hidden layers progressively disentangle the class information from the ECG signals, making the representation of the final layer very discriminative. Unlike the conventional handcrafted features, the feature extraction process in the DNN is purely data-driven, without any expert knowledge.

B. Performance of End-to-End ECG Classification

We applied the aligned feature vectors $\mathbf{x}_i$'s as described in Section II to train a DNN. We set $D = 417$ for all vectors, i.e., the DNN has 417 inputs and 5 output nodes, each output node corresponds to one class in Table III. We used sigmoid nonlinearity in the hidden layers. Stochastic mini-batch (batch size of 128) gradient descent was used in the backpropagation fine-tuning. The learning rate, momentum and maximum number of iterations were set to 0.001, 0.5 and 50, respectively. The DNN has three hidden layers with a structure 417–100–100–100–5. Four experiments were conducted to evaluate the end-to-end approach.

1) Experiment 1 (Exp. 1): In the first experiment, 22-fold cross validation was applied to DS1. Table IV compares the performance of [7] with that of the end-to-end DNN classifier. Note that the proportion of Classes F and Q in the dataset is very small (less than 1%). Thus, the classification performance on these two classes has insignificant contribution to the overall performance. On the other hand, the proportion of Classes S and V is much higher (about 10%) and these two classes contain the majority of arrhythmias. Therefore, we focused on these two classes. To improve the classification performance of Classes S and V, Chazal *et al*. [7] investigated different combinations of feature sets. For simplicity, their best results are shown in Table IV. As can be seen, the overall



(a) 417-Dim. feature vectors (b) Outputs from the 1st hidden-layer

(c) Outputs from the 2nd hidden-layer (d) Outputs from the 3rd hidden-layer

Fig. 6: t-SNE plots of input feature vectors and hidden-layer outputs

TABLE IV: Performance of the classifiers in [7] and our end-to-end classifier (Exp. 1)

		Chazal *et al.* [7]	Proposed method
Overall accuracy		84.5%	93.1%
Class S (S $vs.$ non-S)	SEN	53.3%	69.7%
	SPC	86.7%	86.7%
Class V (V $vs.$ non-V)	SEN	67.7%	88.8%
	SPC	86.7%	86.7%

accuracy of the end-to-end DNN is much higher than that of [7]. In particular, at the same specificity, our DNN achieves a much higher sensitivity for both Class S and Class V.

The MCCs of Classes N, S, V, F and Q achieved by the end-to-end classifier are 0.67, 0.26, 0.67, 0.01 and 0, respectively. To obtain a more balanced MCC performance of the end-to-end classifier, a constant (δ) was added to the output nodes corresponding to Classes S and F so that the classifier has a higher chance of correctly classifying the instances of Classes S and F. Through cross validation on DS1, we found that $\delta = 0.997$ can increase the MCC of Class F from 0.01 to 0.20 without significantly sacrificing the performance of the other classes. More precisely, when $\delta = 0.997$ was added to the outputs of Classes S and F, the MCCs of Classes N, S, V, F and Q become 0.59, 0.34, 0.51, 0.20 and 0, respectively.

Note that this paper does not aim to optimize the solution to solve the data-imbalance problem [36], which is a branch of machine learning research. Therefore, we propose the above simple solution to handle this issue so that the classifier has a greater chance of correctly classifying the instances of the minority classes. The goal is to obtain more balanced MCC values for performance comparison across the five classes. Through the performance investigation, we find that the MCC performance changed to become acceptable. Thus, the above simple solution is sufficient in this work. Actually, we had oversampled the minority classes to deal with this data-imbalance issue. Specifically, we randomly duplicated samples in the minority classes to ensure that the number of instances of each class was balanced in each mini-batch. However, the results were poorer than our current approach.

Fig. 7 shows the ROC curves of the end-to-end classifier for Class S and Class V. Also shown are the operating points (the red $\times$) of the best performing classifier in [7]. Fig. 7 clearly shows that the sensitivity-specificity points in [7] are below the ROC curves of our DNN, suggesting that with a certain range of decision thresholds our DNN achieves better performance (in term of both sensitivity and specificity) than the classifier in [7].

In this experiment, the accuracy of Record 203 was very low (55.9%, the worst case). This record is special in MIT-BIH in that it has the following note [5]:

> "The PVCs are multiform. There are QRS morphology changes in the upper channel due to axis shifts. There is considerable noise in both channels, including muscle artefact and baseline shifts. This is a very difficult record, even for humans."

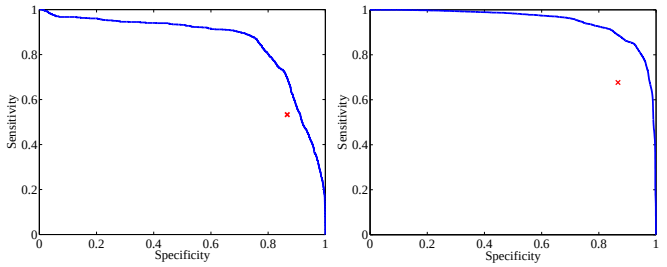


Fig. 7: ROC curves (SEN vs. SPC) of the end-to-end classifier in Exp. 1. *Left panel*: Class S vs. non S (AUC = 0.858). *Right panel*: Class V vs. non V (AUC = 0.951). Red markers correspond to the best performance in [7]. *AUC*: Area under the ROC curve [37].

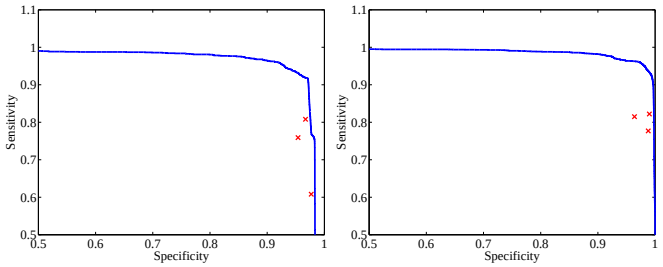


Fig. 8: ROC curves (SEN vs. SPC) of the end-to-end classifier in Exp. 2. *Left panel*: Class S vs. non S (AUC = 0.970). *Right panel*: Class V vs. non V (AUC = 0.991). The red crosses correspond to the best performance in [7], [10], [17].

TABLE V: Performance of the classifiers in [7], [10], [17] and our end-to-end classifier (Exp. 2)

		[7]	[10]	[17]	Proposed
Overall ACC		85.9%	86.4%	89.9%	94.7%
Class S (S *vs.* non-S)	SEN	75.9%	60.8%	80.8%	77.3%
	SPC	95.4%	97.7%	96.7%	97.7%
Class V (V *vs.* non-V)	SEN	77.7%	81.5%	82.2%	93.7%
	SPC	98.8%	96.4%	99.0%	98.8%

TABLE VI: MCC performance of the classifiers in [7], [10], [17] and our end-to-end classifier (Exp. 2)

Method	Class					OMCC
	N	S	V	F	Q	
Chazal *et al.* [7]	0.61	0.52	0.78	0.26	0	0.82
Ye *et al.* [10]	0.57	0.54	0.68	0.05	0	0.83
Raj *et al.* [17]	0.69	0.61	0.82	0.33	0	0.87
Proposed	0.69	0.67	0.91	0.22	0	0.88

We suspect that the special characteristics of the heartbeats in this record are not well represented by the training data (in the other 21 records). To further investigate this issue, we randomly selected 10% of the instances (299 instances) in each class of Record 203, and then added them to the training set. Note that the original number of training instances was 47,999, which is much larger than 299. We did the experiment again, and the classification accuracy increased from 55.9% to 89.3%.

2) Experiment 2 (Exp. 2): In the second experiment, DS1 and DS2 were used as the training set and test set, respectively. Table V shows the performance of the end-to-end classifier and the best results in [7], [10] and [17]. Similar to the results in Exp. 1, the overall accuracy of our approach (End-to-End in Table V) is much higher than that of [7], [10] and [17]. The end-to-end DNN not only achieves a much higher overall accuracy than that of [7], [10] and [17], it also yields a higher sensitivity and specificity for Class S and Class V. Fig. 8 shows the ROC curves of the end-to-end classifier in this experiment. It shows that the best performance in [7], [10] and [17] are below the ROCs of our DNN, which suggests that the end-to-end approach is very promising.

Table VI shows the MCC performance of the classifiers in [7], [10], [17] and our end-to-end classifier. OMCC in the table refers to overall MCC of the five classes. For the end-to-end classifier, the MCCs were obtained by adding the constant ($\delta = 0.997$) found in Exp. 1 to the outputs of Classes S and F. More precisely, we applied cross validation on the training set (DS1) to find an appropriate value for boosting the outputs of Classes S and F to balance the MCCs across the five classes. The results show that the MCC performance of the end-to-end classifier is much better than that in [7] and [10]. Good performance is not only found in Classes N, S, and V, but also in the overall. Compared with [17], our MCC performance is still better except for Class F.

3) Binary Classification: Jun *et al.* [11] proposed using a 6-hidden-layer DNN for PVC beat detection based on the MIT-BIH arrhythmia database. This is a two-class problem in which normal and PVC (NOR and PVC in Table II) heartbeats were extracted for evaluation. In contrast to our raw signal extraction, six handcrafted features were used to represent a heartbeat including R-peak amplitude, RR interval, QRS duration, ventricular activation time, Q-peak amplitude and S-peak amplitude. K-fold cross validation was used to evaluate performance and the performance is optimal when K equals 8. Note that, although 8-fold cross validation was applied in [11], the heartbeats in the cross validation training set and test set could belong to the same patient. However, our previous experiment (Exp. 1) is based on leave-one-subject/patient-out cross validation.

In our experiment, $81,379$ heartbeats were retrieved from the dataset, including $74,478$ normal heartbeats and $6,901$ PVC heartbeats. To make a fair comparison, we also performed 8-fold cross validations. The DNN has the same structure (417–100–100–100–2) as before except for the number of output nodes. Table VII shows the best performance of the classifier in [11] and our end-to-end DNNs. Although the overall accuracy in [11] is high (99.41%), ours (99.70%) is 0.29% higher. Moreover, at very high specificity (99.89%), the sensitivity of the proposed method for Class PVC is still higher than in [11]. Compared with the five-class classification in the previous subsection, this two-class problem is much easier. Not only is the overall accuracy close to 100%, but good performance of detecting PVC beats can also be obtained.

4) Patient-Independent vs. Patient-Specific ECG Classification Systems: Table VIII shows how the patient-specific ECG classification systems in [9], [12], [13] and our patient-independent end-to-end ECG classification system performed. We followed the experimental protocols in [9], [12] and [13]. For the patient-specific classifiers with expert intervention [9],

TABLE VIII: Performance comparisons between the patient-specific classification systems in [9], [12], [13] and our patient-independent classification system on seen patients and unseen patients.

		Test on the remaining 25 min. ECG of 24 seen patients			Test on 22 unseen patients	
		Patient-specific classifiers with expert intervention (Mode 1)		Proposed method	Patient-specific classifiers without expert intervention (Mode 2)	Proposed method
		Ince *et al.* [9]	Kiranyaz *et al.* [12]		Ye *et al.* [13]	
Class S (S $vs.$ non-S)	SEN	62.1%	64.6%	66.2%	61.4%	61.4%
	SPC	98.5%	98.6%	98.6%	99.8%	98.3%
Class V (V $vs.$ non-V)	SEN	83.4%	95.0%	90.5%	91.8%	91.8%
	PC	98.1%	98.1%	98.1%	99.9%	99.5%

TABLE VII: Performance of the classifiers in [11] and our end-to-end classifier

	Jun *et al.* [11]	Proposed method
Overall accuracy	99.41%	99.70%
SEN of class PVC	96.08%	97.68%
SPC of class PVC	Did not specify	99.89%

[12] (Mode 1), to be as fair as possible, we used the first 5 minutes of ECG records (Record No.: 200–234) of 24 patients for training our patient-independent classifier. To evaluate the performance of the classifiers on "seen" patients, we used the remaining 25 minutes of ECG signals of these 24 patients for testing. Note that we used 5 minutes of ECG signals of 24 patients to train a patient-independent classifier. For the patient-specific classifiers without expert intervention [13] (Mode 2), to evaluate the performance of the classifiers on "unseen" patients, we trained a patient-independent classifier based on the ECG records of 22 patients in DS1, and tested the classifier on the other 22 patients in DS2. Table VIII shows that despite patient independency, our patient-independent classifiers achieve comparable performance with the patient-specific classifiers in [9], [12], [13], as evident in the fifth and seventh columns in the table. Bear in mind that any patient-specific classifier requires some patient-specific data or an expensive annotation process for each new patient, therefore our patient-independent classifier definitely has advantages.

C. Summary

The following are advantages of the proposed end-to-end ECG classification method:

1) By using raw-signal extraction and DNNs, the classification performance of our end-to-end system was found to be much better than existing patient-independent systems in terms of sensitivity-vs-specificity ROC and Mathews correlation coefficients; besides that, without expert intervention, its performance is still comparable to patient-specific systems.
2) The end-to-end DNN can perform feature extraction and classification at the same time. Traditional feature extraction methods are limited by the professional knowledge of medical doctors. The end-to-end DNN can overcome such limitation by using aligned raw ECG waveforms as input so that better representations can be obtained for classification.

Note that the classification performance of the proposed method may not be much better than patient-specific classifiers because patient-specific classifiers have patient-specific data, which may be helpful in machine learning. However, the proposed algorithm is a patient-independent classifier which is universal, and it does not need patient-specific data and expert intervention for new patients.

V. Conclusion

This paper introduced an end-to-end ECG classification system. One end of the system receives raw ECG signals and the other end gives beat-by-beat classification decisions. A new preprocessing method, which involves heartbeat segmentation and heartbeat alignment, was proposed to facilitate a deep neural network to form optimal representation of ECG signals and for the classification of heartbeat types.

Four experiments based on the MIT-BIH arrhythmia database were conducted. In the first experiment, 22-fold cross validations on a dataset comprising 50,977 heartbeats and five arrhythmia classes suggest that at the same specificity, the sensitivities of the end-to-end method for Class S and Class V are 16.4% and 21.1% higher (absolute) than those achieved by a conventional method. For all of the five classes, the proposed method achieved higher MCCs and its ROC curves were above the operating points reported in the literature. In the second experiment, the proposed end-to-end DNN was trained on 50,977 heartbeats from 22 patients and tested on 49,668 heartbeats from another 22 patients. Results demonstrated that this end-to-end DNN can capture useful information from the raw ECG signals, enabling it to outperform state-of-the-art arrhythmia classifiers (using either SVM or DNN) that rely on handcrafted ECG features. The third experiment showed the excellent performance (AUC = 0.999) of the proposed method in dealing with the binary ECG classification.

The fourth experiment compared our patient-independent end-to-end ECG classification system with patient-specific ECG classification systems. The results demonstrated that the patient-independent DNN-based classifier generalizes very well to new/unseen patients. The effect of the proposed raw signal extraction method (including segmentation and alignment of complete heartbeats) is remarkable. Thus, the end-to-end ECG classification approach not only outperforms the existing patient-independent classification system, but also performs as well as the patient-specific classification systems.

After using more data to train the patient-independent classifier and testing with more patients, the proposed end-to-end

(input: raw ECG signals; output: beat-by-beat classification decisions) ECG classification system can be introduced as a tool to assist clinicians in diagnosing arrhythmias.

In future we will use long short-term memory (LSTM) [38] instead of DNN classifiers because ECG signals are time series data and LSTM is capable of learning long-term dependencies.

REFERENCES

[1] A. B. de Luna, P. Coumel, and J. F. Leclercq, "Ambulatory sudden cardiac death: Mechanisms of production of fatal arrhythmia on the basis of data from 157 cases," *American Heart Journal*, vol. 117, no. 1, pp. 151–159, 1989.

[2] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: A new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.

[3] F. Agrafioti and D. Hatzinakos, "Fusion of ECG sources for human identification," in *3rd International Symposium on Communications, Control and Signal Processing*, 2008, pp. 1542–1547.

[4] H. Monitor, "Texas heart institute (cited at may 2010)," http://www.texasheart.org/HIC/Topics/Diag/diholt.cfm.

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

[6] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 891–900, 1997.

[7] P. D. Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, July 2004.

[8] S. Osowski, L. T. Hoa, and T. Markiewic, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 582–589, 2004.

[9] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1415–1426, 2009.

[10] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012.

[11] T. J. Jun, H. J. Park, and Y. H. Kim, "Premature ventricular contraction beat detection with deep neural networks," in *15th IEEE International Conference on Machine Learning and Applications*, 2016, pp. 859–864.

[12] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.

[13] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "An automatic subject-adaptable heartbeat classifier based on multiview learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 6, pp. 1482–1492, Nov. 2016.

[14] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information sciences*, vol. 405, pp. 81–90, 2017.

[15] J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, O. S. Lih, M. Adam, R. S. Tan, M. Chen, and U. R. Acharya, "Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals," *Computers in Biology and Medicine*, vol. 94, pp. 19–26, 2018.

[16] U. R. Acharya, H. Fujita, O. S. Lih, 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, pp. 190–198, 2017.

[17] S. Raj and K. C. Ray, "Sparse representation of ECG signals for automated recognition of cardiac arrhythmias," *Expert Systems with Applications*, vol. 105, pp. 49–64, 2018.

[18] ANSI/AAMI, "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," *Association for the Advancement of Medical Instrumentation*, no. EC57, 1998.

[19] A. L. G. *et al.*, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.

[20] P. M. Rautaharju, B. Surawicz, and L. S. Gettes, "AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram," *Journal of the American College of Cardiology*, vol. 53, no. 11, pp. 982–991, 2009.

[21] J. S. S. *et al.*, "Value of the P-wave signal-averaged ECG for predicting atrial fibrillation after cardiac surgery," *Circulation*, vol. 88, no. 6, pp. 2618–2622, 1993.

[22] D. M. B. *et al.*, "Microvolt T-wave alternans and the risk of death or sustained ventricular arrhythmias in patients with left ventricular dysfunction," *Journal of the American College of Cardiology*, vol. 47, no. 2, pp. 456–463, 2006.

[23] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," *Advances in Neural Information Processing Systems*, vol. 19, p. 153, 2007.

[24] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[25] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning-a new frontier in artificial intelligence research," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, Nov. 2010.

[26] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[27] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *13th International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[28] D. Erhan, Y. Bengio, A. Courville, P. A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *The Journal of Machine Learning Research*, vol. 11, pp. 625–660, 2010.

[29] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, Mar. 1985.

[30] E. J. Boyko, "Ruling out or ruling in disease with the most sensitiue or specific diagnostic test: Short cut or wrong turn?" *Medical Decision Making*, vol. 14, no. 2, pp. 175–179, 1994.

[31] A. Laupacis and N. Sekar, "Clinical prediction rules: a review and suggested modifications of methodological standards," *Jama*, vol. 277, no. 6, pp. 488–494, 1997.

[32] J. A. Swets, *Signal detection theory and ROC analysis in psychology and diagnostics : collected papers*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[33] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochimica et Biophysica Acta (BBA)-Protein Structure*, vol. 405, no. 2, pp. 442–451, Oct. 1985.

[34] M. W. Mak, J. Guo, and S. Y. Kung, "PairProSVM: protein subcellular localization based on local pairwise profile alignment and SVM," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 5, no. 3, pp. 416–422, 2008.

[35] L. V. D. Maaten and G. E. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[36] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429–449, 2002.

[37] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.