Subject-Specific Detection of Ventricular Tachycardia Using Convolutional Neural Networks

B S Chandra, C S Sastry, S Jana

Indian Institute of Technology Hyderabad, Telangana, India

Abstract

Onset of ventricular tachycardia (VT) is clinically significant, including as a trigger to defibrillator implants. In this paper, we propose a reliable technique to detect such onset using convolutional neural networks (CNNs). The proposed CNN adds convolution and pooling layers below the input layer and above the hidden and output layers of usual neural network (NN). Such layers would learn suitable linear features from training data, while eliminating the need to extract the traditionally used adhoc features. Employing such subject-specific features, we reported the performance of the proposed classifier using Creighton University ventricular tachyarrhythmia database (CUVT). In particular, we achieved mean (± standard deviation) performance of 95.6 (± 00.6) using subject-specific evaluation scheme over 100 random independent iterations.

1. Introduction

Cardiovascular diseases (CVDs) are a leading cause of death worldwide [1]. Accordingly, their management sets an important healthcare objective. An indispensable tool in diagnosing and monitoring CVDs is electrocardiogram (ECG). In certain scenarios, ECG from the patient is continuously monitored to detect various arrhythmic conditions. Specifically, in home based monitoring of high-risk patients, automatic detection of ventricular tachycardia (VT) can trigger the defibrillator implants and correct such life-threatening arrhythmia. Indeed timely and accurate detection of VT assumes significance in critical CVD management.

Numerous algorithms have been reported for VT detection based on various features extracted from time domain, transform domain and a mixture of features from both time and transform domains [2–4]. Though such algorithms have reported high classification performance, majority of them remain unreliable to be used in practice. Traditionally, a fixed set of hand-crafted features are used in the classifier design. However, ECG has high inter-patient variations, and such generic features may not adequately represent the underlying characteristic of the signal specific to the individuals. In addition, the performance reported by various algorithms in the literature lack adherence to Association for the Advancement of Medical Instrumentation (AAMI) standards [5], making it difficult to compare with other algorithm and extend for practical implementation. Against this backdrop, we propose a novel subject-specific VT detection scheme using convolutional neural networks (CNNs), and demonstrated its utility while adhering to AAMI recommended practice.

The proposed CNN adds convolution and pooling layers below the input layer and above the hidden and output layers of usual neural network (NN) [6, 7]. While traditional methods, such as NN and support vector machine (SVM), operate on ad hoc features, our additional layers learn suitable linear features from training data, providing an advantage. We demonstrated the efficacy of the proposed approach using 35 patients data from Creighton University ventricular tachyarrhythmia database (CUVT) database [8]. Specifically, we performed three experiments, where the respective training data consisted of *subject-oblivious* a random 80% of all signal vectors; *unseen subject* signal vectors from a random 28 subjects (80%); and *subject-specific* signal vectors from a random 28 subjects and suitable 20% duration of rest of the subjects. In each case, remaining vectors were used for testing. Each experiment underwent 100 independent trials, our system achieved a mean (± standard deviation) classification accuracy of 95.6% (± 0.06) with subject-specific evaluation.

The contribution of the present work is summarized as follows: (i) adopted a subject-specific approach for VT detection, and compared with subject-oblivious (optimistic) and unseen subject (conservative) experiments; (ii) learned features using CNN instead of using ad hoc ones; and (iii) achieved high mean performance as well as high robustness (standard deviations two orders of magnitude lower compared to existing methods).

The paper is organized as follows. In Section 2 we present the methodology. In Section 3 we present the results. Finally in Section 4 we conclude the paper with a discussion.

2. Methodology

In this section, we first present a mathematical formalism for ECG classification, then present the CNN based classifier and finally outline the evaluation strategy for the proposed classifier.

2.1. Formalism

A desired classifier specifies two mutually exclusive and exhaustive subsets Γ_1 and Γ_2 of set Γ of possible ECG signal x as follows. Any beat $x \in \Gamma_1$ is declared normal, while any signal $x \in \Gamma_2$ is declared a VT. Our goal is to find the subsets Γ_1 and Γ_2 such that sensitivity Se (fraction of VT signals correctly detected as VT) and specificity Sp (fraction of normal signals correctly detected as normal) of the underlying classifier are maximized.

2.2. Convolutional neural networks

In this paper, we propose a novel ECG classification approach based on 1-D convolutional neural networks (CNNs). CNNs adds convolution and pooling layers below the input layer and above the hidden and output layers of usual neural networks [6, 7]. A sample CNN architecture is illustrated in Figure 1 with a single convolution and pooling layers followed by the the fully connected network to output layer. Convolutional layer will have k filters (or kernels) of size q, where q is smaller than the length of the input signal vector m. Each filter is convolved with the signal vector, and an additive bias followed by nonlinearity is applied to produce k feature maps of size $m - q + 1$. The forward propagation operation in convolutional layer (from layer $l - 1$ to layer l) can be expressed as follows

$$z_k^{(l)} = b_k^{(l-1)} + \sum_{i=1}^{N_{l-1}} \text{conv}(w_k^{(l-1)}, a^{(l-1)}), \quad (1)$$

$$a_k^{(l)} = f(z_k^{(l)}), \quad f(z) = \frac{1}{1 + \exp(-z)}, \quad (2)$$

where conv is the convolution operation without zero padding on the boundaries, $a^{(l-1)}$ is the input to the $(l-1)^{th}$ layer (a^1 being the input signal vector). b_k^{l-1} and $w_k^{(l-1)}$ are the bias and weights of the kth filter at layer $l - 1$ respectively. The function f is the non-linear activation function called the sigmoid or logistic function.

After convolution layer, we decide the size of the pooling region, say p to pool our convolved features. To this end, we divide our convolved features into disjoint regions of length p, and take the mean (or maximum) feature activation over these regions to obtain the pooled convolved features. These pooled features from multiple filters are
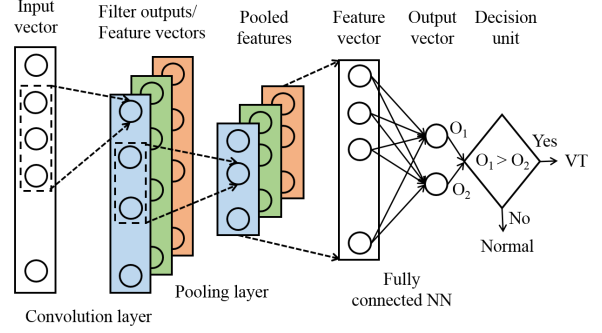

Figure 1. CNN architecture.

then stacked together to form a single feature vector. Following the pooling layer there may be any number of fully connected layers. The densely connected layers are identical to the layers in a standard multilayer neural network. Here the forward and the back propagation equations remain unaltered from that of the traditional neural networks.

For the binary classfication task (noram versus VT), we denote the output of the final layer to be O_1 and O_2. We interpret O_1 and O_2 as the probability of occurrence of each class (i.e., $O_1 + O_2 = 1$). Finally, given a labeled training data with $y^{(i)} = 0$ and 1 representing the labels of normal and VT respectively, for the ith signal vector. The cost function at output layer is given by

$$J = -\sum_i (y^{(i)} \log(O_1^{(i)}) + (1 - y^{(i)}) \log(1 - O_1^{(i)})). \quad (3)$$

Our goal is to minimize J as a function of W and b. To this end, we train our neural network by initializing each parameter $W_k^{(l)}$ and each $b_i^{(l)}$ to a small random value and optimize using batch gradient descent. We now describe the back propagation algorithm for convolutional and pooling layers, which gives an efficient way to compute the partial derivatives that are used in gradient descent update equations. Denoting the error at layer $l + 1$ by $\delta^{(l+1)}$, if the lth layer is a convolution and pooling layer, error is backpropagated as

$$\delta_k^{(l)} = \text{upsample}\left((W_k^{(l)})^T \delta_k^{(l+1)}\right) f'(z_k^{(l)}), \quad (4)$$

where k indexes the filter number and $f'(z_k^{(l)})$ is the derivative of the activation function. The upsample operation propagates the error through the pooling layer by calculating the error with respect to each unit incoming to the pooling layer. For mean pooling, upsample operation uniformly distributes the error from a single pooling unit to all the units that feed into it from the previous layer. Whereas in max pooling, the unit which was chosen as the max receives all the error. Finally, gradients for filter

weights are obtained by convolving the error vector $\delta_k^{(l)}$ with the activations corresponding to the filter k at layer l

$$\nabla_{W_k^{(l)}} J = \sum_{i=1} \text{conv}(a_i^{(l)}, \delta_k^{(l+1)}), \tag{5}$$

$$\nabla_{b_k^{(l)}} J = \sum (\delta_k^{(l+1)}). \tag{6}$$

With the convolutional network and gradient descent update, the parameters are learnt based on the training data. Given test signal, signal is forward propagated and output probabilities O_1 and O_2 are obtained. Finally, signal is assigned to normal class if O_1 is greater than O_2 and VT otherwise. Having equipped with the CNN classifier, we now present the evaluation strategy to validate the classifier performance.

2.3. Evaluation strategy

Traditionally, partitioning of database into training and test sets is performed either in a subject-oblivious or unseen-subject manner. In the former, ECG segments were extracted from all or a few of the records were clustered, based on the rhythm label. A certain fraction data from each cluster was selected as the training set and rest of the data as test set. In such scheme, same subject data may be represented in both training and test sets, resulting in overly optimistic performance estimates. In contrast, the latter seeks to account for inter-subject variability, and constitutes training and test sets with beats from distinct subsets of records, leading to an overly conservative estimate of performance.

A hybrid scheme called subject-specific training has been recommended in AAMI standards, in which a subject-oriented approach is taken with the following modification. A few subject-specific data are added to the training set. Such subject-specific approach often provides a reasonable performance estimate, which is less optimistic than the performance estimated using subject-oblivious scheme, and less conservative than unseen-subject approach. Accordingly, we adopt a subject-specific paradigm in the preset work. Further to represent a more practical performance, we reported the performance averaged over 100 independent random trials.

3. Experiments and results

Now, we turn to evaluating the proposed system. To begin, we describe the database under consideration and the preprocessing steps.
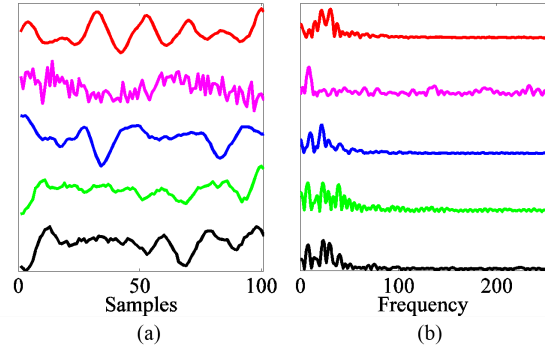

Figure 2. Leraned filters in (a) time domain; (b) frequency response

3.1. Database and Preprocessing

We validated the proposed method using Creighton University ventricular tachyarrhythmia (CUVT) database available from the PhysioBank archives [8]. The database consists of 35 eight-minute ECG recordings digitized at 250 Hz with 12-bit resolution over a 10V range. Each record is annotated with the onset and offset of VT episodes. We segmented entire record into overlapping signal vectors, each of duration five seconds with an overlap of 0.2 seconds, amounted to 22,772 such vectors.

3.2. Learned features

We used a single convolution layer and a pooling layer followed by logistic regression for classification. We used 5 filter kernels with length 101 and a pool dimension of 2. As alluded earlier, the proposed scheme learns linear features to perform the classification. Figure 2 illustrates the kernels learned and their frequency response. Note that, each filter passes only certain frequencies that are optimized for classification performance on training set. The feature vector generated after convolving each of the filter with both normal and VT signal is shown in Figure 3. Clearly R-peak values are enhanced in various filters for normal signals. While for VT signals with no clear R-peak, each filter has a response that distinguishes VT signal from its rival class.

3.3. Classification performance

As mentioned earlier, we evaluated our system using three schemes of partitioning of database into training and test sets. (i) Subject-oblivious scheme selects a random 80% of all signal vectors for training and rest of the data for testing. (ii) Unseen-subject selects a random 28 subjects (80%) data for training and rest of the subjects data are used for testing. (iii) Subject-specific scheme selects
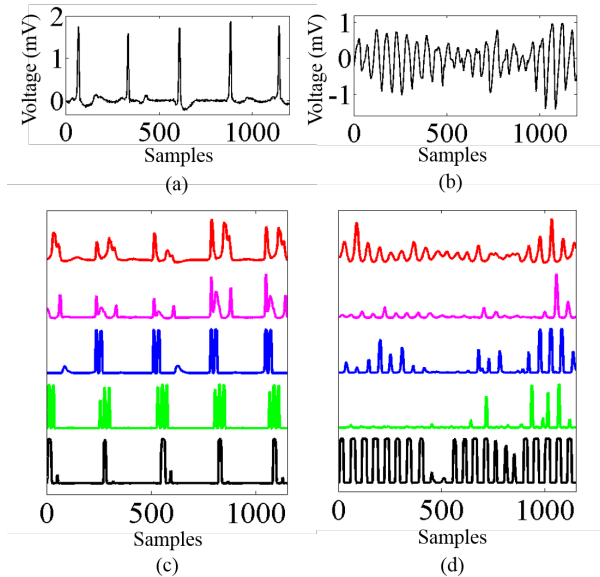
Figure 3. (a) Typical normal signal; (b) Typical VT signal; Output of the convolution layer (features) corresponding to five filters for (c) normal signal and (d) VT signal

Method	Se (%)	Sp (%)	Acc (%)
Subject-oblivious	98.3±0.06	98.5±0.03	98.4±0.04
Unseen subject	94.4±0.07	89.2±0.1	92±0.06
Subject-specific	**95.6±0.06**	**96.6±0.03**	**95.9±0.03**

Table 1. Classification performance

signal vectors from a random 28 subjects and first 20% of the signal vectors (both normal and VT) from remaining 7 subjects for training. Rest of the data from 7 subjects were used for testing. In each case, we recorded the classifier performance for 100 independent trials, and mean and standard deviation of sensitivity (Se), specificity (Sp) and accuracy (Acc) are reported in the Table 1. Clearly, the mean performance of subject-oblivious scheme is significantly high. However, it remain unrealistic as the same patient data is well represented in both training and test sets. Unseen-subject represent a conservative estimate of performance in view of the subject-data not represented in the training and test sets. With addition of patient-specific data, subject-specific scheme represent realistic performance of the classifier. Direct comparison of performance with other methods may not be valid as the classification performance depends on the data adopted for training and testing of the classifier.

4. Discussion

In this study, we proposed a patient-specific VT detector using CNNs. We optimized weights of the convolutional layer to extract the features specific to the subject that max-

imizes the underlying classifier performance. Our solution not only negates the necessity to extract hand-crafted manual features, but also eliminates the need for any kind of pre- and post-processing of ECG signals, rendering it to be readily useful for real-time applications. We validated the performance of our classifier on the benchmark CUVT database while adhering to AAMI standards. The proposed method yields a mean (standard deviation) sensitivity and specificity of 95.6% (± 0.06) and 96.6% (± 0.03) respectively with subject-specific evaluation. Such evaluation is deemed as a more realistic estimate of potential performance of the proposed method in real applications. As a future work, we intend to test the performance of the system on a larger set of database and design the hardware implementation. Further, considering the criticality of VT detection, classification has to be designed to optimize high sensitivity detection.

References

[1] World Health Organization Cardiovascular diseases (cvds): Fact sheet no. 317, 2011, World Health Organization.

[2] Thakor NV, Zhu YS, Pan KY. Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. IEEE Transactions on Biomedical Engineering 1990; 37(9):837–843.

[3] Owis MI, Abou-Zied AH, Youssef AB, Kadah YM. Study of features based on nonlinear dynamical modeling in ecg arrhythmia detection and classification. IEEE transactions on Biomedical Engineering 2002;49(7):733–736.

[4] Li Q, Rajagopalan C, Clifford GD. Ventricular fibrillation and tachycardia classification using a machine learning approach. IEEE Transactions on Biomedical Engineering 2014;61(6):1607–1613.

[5] ECAR A. Recommended practice for testing and reporting performance results of ventricular arrhythmia detection algorithms. Association for the Advancement of Medical Instrumentation 1987;.

[6] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems. 2012; 1097–1105.

[7] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks 1995;3361(10):1995.

[8] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

Address for correspondence:

B S Chandra,
Dept. of Electrical Engineering, CPS Lab,
Indian Institute of Technology Hyderabad, India-502285,
bschandra@iith.ac.in