K-margin-based Residual-Convolution-Recurrent Neural Network for Atrial Fibrillation Detection

Yuxi Zhou[1,2] , **Shenda Hong**[1,2] , **Junyuan Shang**[1,2] , **Meng Wu**[1,2] , **Qingyun Wang**[1,2] ,
Hongyan Li[1,2,*] and **Junqing Xie**[3]

[1]School of Electronics Engineering and Computer Science, Peking University, Beijing, China
[2]Key Laboratory of Machine Perception (Peking University), Ministry of Education, Beijing, China
[3]Medical Informatics Center, Peking University, Beijing, China
{joy_yuxi, hongshenda, sjy1203, wumeng93, wangqingyun, leehy, xjq}@pku.edu.cn

Abstract

Atrial Fibrillation (AF) is an abnormal heart rhythm which can trigger cardiac arrest and sudden death. Nevertheless, its interpretation is mostly done by medical experts due to high error rates of computerized interpretation. One study found that only about 66% of AF were correctly recognized from noisy ECGs. This is in part due to insufficient training data, class skewness, as well as semantical ambiguities caused by noisy segments in an ECG record. In this paper, we propose a K-margin-based Residual-Convolution-Recurrent neural network (K-margin-based RCR-net) for AF detection from noisy ECGs. In detail, a skewness-driven dynamic augmentation method is employed to handle the problems of data inadequacy and class imbalance. A novel RCR-net is proposed to automatically extract both long-term rhythm-level and local heartbeat-level characters. Finally, we present a K-margin-based diagnosis model to automatically focus on the most important parts of an ECG record and handle noise by naturally exploiting expected consistency among the segments associated for each record. The experimental results demonstrate that the proposed method with 0.8125 F_{1NAOP} score outperforms all state-of-the-art deep learning methods for AF detection task by 6.8%.

1 Introduction

Atrial Fibrillation (AF) is an abnormal heart rhythm characterized by rapid and irregular beating of the atria, which can trigger cardiac arrest and sudden death as a consequence of pumping blood less effectively [Ye *et al.*, 2010; Kass and Clancy, 2005]. The Electrocardiograph (ECG) has been a cornerstone for the detection and diagnosis of such condition for a long time. However, its interpretation is mostly done by medical experts due to the high error rates of computerized interpretation. It has been found that good performance was shown on carefully-selected often clean data [Clifford *et al.*, 2017], and only about 66% of AF predictions were correctly recognized from noisy ECG data [Shah
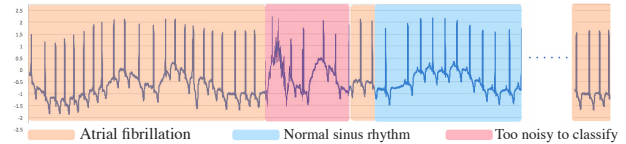
---

*Contact Author

Figure 1: An example of an AF record with noisy segments.

and Rubin, 2007]. This is especially problematic in developing countries, where the availability of clinics and medical experts is low [Gradl *et al.*, 2012] [Silva *et al.*, 2011]. Accordingly, there is a definite need for automatic, accurate physiological monitoring solutions for AF detection (without domain knowledge and features given by domain experts in advance), which can be used in home or ambulatory settings.

Despite the significance of this problem, it is challenging to reliably detect AF from noisy ECGs due to the following aspects. First, there is insufficient training data for AF detection, since well-labelled ECG recordings are hard to acquire. Second, the inevitable existence of class skewness in ECGs (as diseases happen rarely) may result in poor performance. Then, it is hard to capture both local heartbeat-level characters and long-term rhythm-level trend in an ECG record. Moreover, there may be semantical ambiguities caused by noisy segments in an ECG record as shown in Figure 1. Learning features from these noisy segments will lead to poor Deep Neural Network (DNN) models.

In this paper, we propose a K-margin-based Residual-Convolution-Recurrent neural network (K-margin-based RCR-net) for AF detection from noisy ECGs. We first employe a skewness-driven dynamic augmentation method to handle problems of data inadequacy, class imbalance and high computing cost in ECG analysis. Then, we propose a multi-view learning method based on a novel Residual-Convolution-Recurrent neural networks (RCR-net) to automatically extract both long-term rhythm-level and local heartbeat-level characters. Besides, a K-margin-based diagnosis model is presented to identify the classification of each ECG by selecting its most probable label of relevant top-K augmented segments with the least certainty margin, which can automatically focus on the most important segments associated with an ECG record and handle noise by naturally exploiting expected consistency among the segments associ-
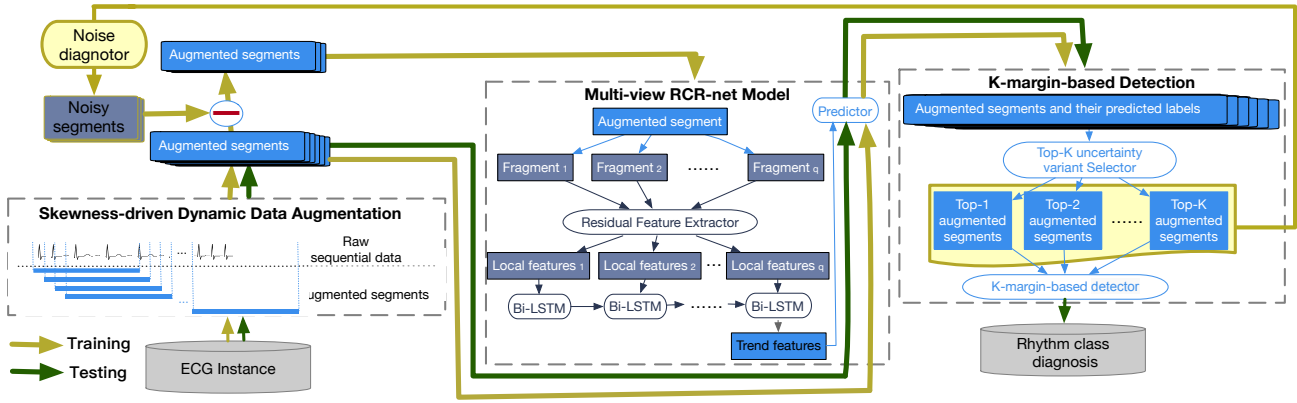
Figure 2: The framework of K-margin-based RCR-net model

ated for each record. We carry out thorough experiments on PhysioNet/Computing in Cardiology Challenge 2017 Dataset [Clifford *et al.*, 2017]. The experimental results demonstrate that the proposed method with 0.8125 F_{1NAOP} score outperforms all state-of-the-art deep learning methods for AF detection task by 6.8%

2 Related Work

AF detectors can be divided into two categories including feature engineered methods and deep learning based methods.

The feature engineered methods are highly related to cardiological knowledge which can achieve high accuracy if the recorded ECG signals are clean and high-resolution. In AF detection task, a process of using cardiological knowledge to extract informative values is essential to make classifiers able to distinguish AF patients from the population. Although engineering effective features is difficult and time-consuming, there are many excellent works that proposed various kinds of features and got good detection performance [Tateno and Glass, 2001; Linker, 2016; Carrara *et al.*, 2015]. These methods help cardiologist to promote medical service quality in hospital. However, the biggest problem is that it's hard to extract accurate features from noisy contaminated ECGs.

Recently, deep learning-based methods like Convolutional Neural Network (CNN) [Kiranyaz *et al.*, 2015; Xiong *et al.*, 2017; Andreotti *et al.*, 2017; Chandra *et al.*, 2017; Kamaleswaran *et al.*, 2018; Hannun *et al.*, 2019], Recurrent Neural Network (RNN) [Schwab *et al.*, 2017], Convolutional Recurrent Neural Network (CRNN) [Zihlmann *et al.*, 2017], and combined methods [Hong *et al.*, 2017; Hong *et al.*, 2019] have achieved success in AF detection due to their powerful automatic feature learning ability. However, as far as we are concerned, these methods still have high error rates for ECG diagnosis when encountering insufficient training data, class skewness, as well as the semantic ambiguities of noisy ECGs.

3 Methods

3.1 Problem Definition and General Framework

AF detection is the task of automatically classifying an ECG into one of cardiac arrhythmia classes. Formally,

we denote the training dataset as $\mathcal{D} = \{X, Y\}$ where $X = [\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots, \boldsymbol{x}^{(N)}]$ are ECG sequence inputs, $Y = [y^{(1)}, y^{(2)}, \dots, y^{(N)}]$ the corresponding label set and N the total number of training data. Given the labelled training dataset $\mathcal{D}$, our goal is to learn a predictive model which takes an unlabelled ECG sequence $\boldsymbol{x}^{(i)}$ as input and outputs the prediction $\hat{y}^{(i)} \in \mathcal{C}$ where $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$ is a set of m different rhythm classes.

In this paper, we consider a DNN model as the basic sequence classifier. Recent deep learning works have demonstrated significant success on sequential classification tasks, that can naturally integrate and extract hierarchy features automatically. For the purpose of benefiting from both automatic feature extraction from CNN and capturing the long-term trend from RNN, we employ a Residual-Convolutional-Recurrent Neural Network (RCR-net) as shown in Figure 2.

Nevertheless, a RCR-net model still can't be directly applied for reliably detecting AF from noisy ECG segments. On the one hand, ECGs are generated continuously with hundreds of millions of points of each patient, training a DNN model on such a long time period may result in high computing complexity. On the other hand, the class skewness and semantical ambiguities caused by noisy ECG segments may lead to poor, or even unacceptable quality of DNN models.

To solve the aforementioned problems, we adapt a K-margin-based learning approach. The framework of K-margin-based RCR-net is shown in Figure 2. In detail, we firstly preprocess raw data using Skewness-driven Dynamic Data Augmentation (Sec. 3.2) to relieve the class imbalance problem. Then, the RCR-net model (Sec. 3.3) is trained using the labelled augmented segments. It is noteworthy that only a portion of the segments in each record participate in the RCR-net learning process. Finally, a K-margin-based diagnosis model (Sec. 3.4) is employed to identify the classification of each ECG record by automatically focusing on the most important segments and handling noise.

3.2 Skewness-driven Dynamic Data Augmentation

As introduced before, to solve or relieve the problems of inadequate data, classes skewness and high computing complexity, we propose a skewness-driven dynamic augmentation
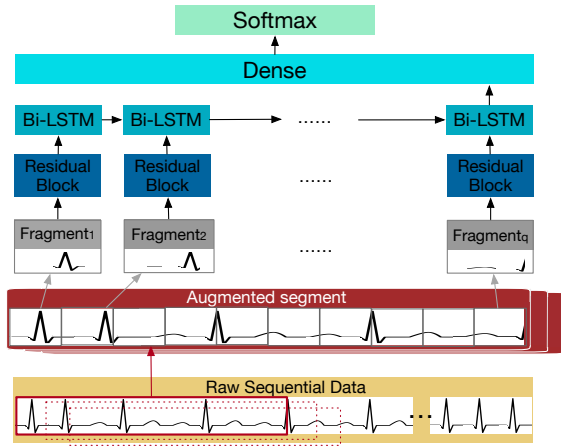
Figure 3: The high-level architecture of RCR-net model

Figure 4: An example of "hard" segments with noisy labels caused by data augmentation process.

method to handle these problems. Specifically, we deploy slide-and-cut methodology to generate more short-term ECG segments based on original long ECG records. Usually, slide-and-cut needs two predefined parameters: 1) *windows_size* (denoted as w): length of cut, 2) *stride* (denoted as s): length of slide. In our dynamic data augmentation process, we employ a skewness-driven dynamic stride. The setting of *stride* gets smaller for records whose labels are scarce, and gets larger for records whose labels are common, constrained by a maximum stride threshold. Formally, given a maximum stride threshold MS and m labels $C = \{c_1, c_2, \ldots, c_m\}$, the stride of record with label c_j is given by:

$$s_{c_j} = \lceil MS \times \frac{|\text{records labelled } c_j|}{\max_{s=1}^{m} |\text{records labelled } c_s|} \rceil \quad (1)$$

Notice that if one ECG record has less than *windows_size* length, we pad it with zeros at the end. Therefore, a T-segments array $\boldsymbol{X}^{(i)} = [\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(i)}, \ldots, \boldsymbol{x}_T^{(i)}]$ can be generated from an ECG record $\boldsymbol{x}^{(i)} \in X$, where $t \in \{1, 2, \ldots, T\}$-th segment $\boldsymbol{x}_t^{(i)}$ is a continuous real-valued vector $[x_{t1}^{(i)}, x_{t2}^{(i)}, \ldots, x_{tw}^{(i)}] \in \mathbb{R}^w$ and w is the window size. The label $y^{(i)}$ of i-th record $\boldsymbol{x}^{(i)}$ is assigned to all the segments in its augmented segments array $\boldsymbol{X}^{(i)}$. For clarity, we denote a segments labels set as $Y^{(i)} = [y_1^{(i)}, y_2^{(i)}, \ldots, y_T^{(i)}]$ for each ECG sequence $\boldsymbol{x}^{(i)}$ where $y_t^{(i)}$ is the corresponding label for t-th segment $\boldsymbol{x}_t^{(i)}$.

3.3 Multi-view Residual-Convolution-Recurrent Neural Network

Recently, DNN methods have achieved great success in AF detection [Wu *et al.*, 2016; Majumdar and Ward, 2017; Kiranyaz *et al.*, 2015; Hannun *et al.*, 2019]. To build a more general deep learning method which could automatically capture anti-noise characteristics for a wider range of arrhythmias, we want to extract and integrate both local heartbeat-level characters and long-term rhythm-level trend from an ECG record, namely multi-view deep features. From this
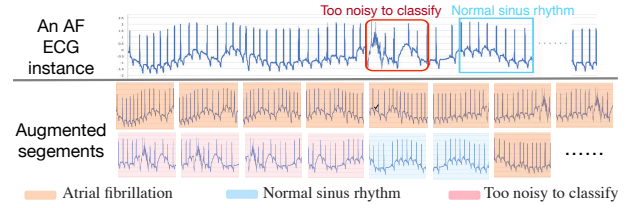
point of view, we design a multi-view residual-convolution-recurrent neural network architecture (RCR-net) to benefit from both the power of automatic feature extraction using CNN and the long-term trend captured using RNN. The high-level architecture of the RCR-net is shown in Figure 3. The network takes augmented ECG segments as input, and outputs the predictions.

In detail, RCR-net consists of a 33-layer stacked residual block [He *et al.*, 2016], 1-layer recurrent block and 1-layer fully connected block. The residual block layers aim to automatically extract more effective local heartbeat-level features by constructing a very deep model via residual connections between blocks. The recurrent layer is designed to capture underlying rhythm-level structure in ECG data. Here, we use Bi-directional Long-Short Term Memory (Bi-LSTM) cells which can capture long term trend by utilizing a gated architecture. Then, the predictions are made by a fully connected layer and a softmax layer. Finally, we compute cross-entropy loss for objective function, and optimize the loss for training the neural network.

3.4 K-margin-based Diagnosis Model

Skewness-driven dynamic data augmentation is essential to boost the model performance by handling the problems of data inadequacy, class imbalance and high computing complexity in ECG analysis. However, as shown in Figure 4, it inevitably generates "hard" ECG segments due to noisy labels. Therefore, to enhance the robustness of RCR-net, we propose to compute cross-entropy objective function of only a portion of the selected ECG segments.

In order to select the proper segments for computing cross-entropy, we first propose a multi-class uncertainty variant measurement called least uncertainty margin. Given a trained RCR-net model, one can get all the predictions of augmented ECG segments for each ECG record. Then, we define the uncertainty margin of t-th segment $\boldsymbol{x}_t^{(i)}$ for the given ECG record $\boldsymbol{x}^{(i)}$ under trained RCR-net model as:

$$Margin(\boldsymbol{x}_t^{(i)}) = P(\ddot{\hat{y}}_t^{(i)}|\boldsymbol{x}_t^{(i)}) - P(\dot{\hat{y}}_t^{(i)}|\boldsymbol{x}_t^{(i)}) \quad (2)$$

where $\dot{\hat{y}}_t^{(i)}$ and $\ddot{\hat{y}}_t^{(i)}$ are the most and second-most probable predicted classes of $\boldsymbol{x}_t^{(i)}$ using the RCR-net model. Intuitively, segment with the least uncertainty margin is the most confident one for the given ECG record under the trained model, because the trained model has little doubt in differentiating between the two most probable classes. By contrast,

---

Algorithm 1 K-LABELS($\boldsymbol{X}^{(i)}, K, \phi$)

Input: A T-segments array $\boldsymbol{X}^{(i)} = [\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(i)}, \ldots, \boldsymbol{x}_T^{(i)}]$ of an input ECG record $\boldsymbol{x}^{(i)}$, trained RCR-net model ϕ
Parameter: An integer K
Output: Top-K most confident segments $\boldsymbol{X}_{1:K}^{(i)}$ and their label predictions $\hat{Y}_{1:K}^{(i)}$

1: $\boldsymbol{X}_{1:K}^{(i)} \leftarrow \emptyset, \ \hat{Y}_{1:K}^{(i)} \leftarrow \emptyset, \ k \leftarrow K$
2: **while** $k > 0$ **do**
3: $\quad \boldsymbol{x}_*^{(i)} \leftarrow$ Most confident segment predicting by $\phi(\boldsymbol{X}^{(i)})$ using Eq. 3
4: $\quad \hat{y}_*^{(i)} \leftarrow$ Most confident label predicting by $\phi(\boldsymbol{X}^{(i)})$ using Eq. 4
5: $\quad \boldsymbol{X}^{(i)} \leftarrow \boldsymbol{X}^{(i)} \setminus \boldsymbol{x}_*^{(i)}$ # Implementation: remove segment index of $\boldsymbol{x}_*^{(i)}$ from segment indexes list of $\boldsymbol{X}^{(i)}$
6: $\quad \boldsymbol{X}_{1:K}^{(i)} \leftarrow \boldsymbol{X}_{1:K}^{(i)} \cup \{\boldsymbol{x}_*^{(i)}\}, \ \hat{Y}_{1:K}^{(i)} \leftarrow \hat{Y}_{1:K}^{(i)} \cup \{y_*^{(i)}\}$
7: $\quad k \leftarrow k - 1$
8: **end while**
9: **return** $\boldsymbol{X}_{1:K}^{(i)}, \ \hat{Y}_{1:K}^{(i)}$

---

segments with larger margins are more ambiguous. Therefore, we define the segment with least uncertainty margin of i-th input ECG record $\boldsymbol{x}^{(i)}$ with T-segments array under trained RCR-net model as:

$$\boldsymbol{x}_*^{(i)} = \underset{\boldsymbol{x}_t^{(i)}, t \in \{1,2,\ldots,T\}}{\arg\min} (Margin(\boldsymbol{x}_t^{(i)})) \tag{3}$$

where $\boldsymbol{x}_t^{(i)}$ is the t-th segment of the input ECG record $\boldsymbol{x}_i$. Least margin diagnosis aims to find the predicted classes of an ECG record by applying the most confident strategy. In a similar way, the most confident label of an input ECG record $\boldsymbol{x}^{(i)}$ under trained RCR-net model is defined as follows:

$$y_*^{(i)} = \underset{\dot{y}_t^{(i)}, t \in \{1,2,\ldots,T\}}{\arg\min} (P(\ddot{y}_t^{(i)}|\boldsymbol{x}_t^{(i)}) - P(\dot{y}_t^{(i)}|\boldsymbol{x}_t^{(i)})) \tag{4}$$

Based on the above discussion, the proposed K-margin-based diagnosis algorithm (K-LABELS) is described in Algorithm 1. It takes an ECG record $\boldsymbol{x}^{(i)}$ as an input, and the top-K most confident segments and their labels under trained model ϕ can be identified by it. Given an ECG record $\boldsymbol{x}^{(i)}$ and its augmented segments array $\boldsymbol{X}^{(i)} = [\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(i)}, \ldots, \boldsymbol{x}_T^{(i)}]$, it requires K iterations where K is the number of most confident segments associated with it in Algorithm 1. The top-1 most confident segment and its label needs to be found out from the segment array in iteration, and then it will be removed from the segment array. Given m rhythm labels $\mathcal{C} = [c_1, c_2, \ldots, c_m]$, after K iterations, the top-K most confident segments $\boldsymbol{X}_{1:K}^{(i)}$ and their label $\hat{Y}_{1:K}^{(i)}$ for the given ECG record $\boldsymbol{x}^{(i)}$ under trained model ϕ can be identified.

To avoid learning features from the noisy segments, it is expected that the model can automatically focus on the most important segments which could represent the features of the given ECG record. Intuitively, the top-K most confident segments $\boldsymbol{X}_{1:K}^{(i)}$ can be used for the training process, instead of the whole segments array $\boldsymbol{X}^{(i)}$. However, as they are selected by the RCR-net model, it is unreliable to use them for training when the model is unreliable. Therefore, for each ECG record $\boldsymbol{X}^{(i)}$, we can compute the average probabilistic prediction of all of its augmented segments:

$$\alpha_i = \frac{1}{T} (\sum_{t=1}^{T} P(\dot{y}_t^{(i)}|\boldsymbol{x}_t^{(i)})) \tag{5}$$

where T is the number of augmented segments within candidate $\boldsymbol{x}^{(i)}$, $P(\dot{y}_t^{(i)}|\boldsymbol{x}_t^{(i)})$ is the prediction probability of t-th segment of $\boldsymbol{x}^{(i)}$. If $\alpha_i > 0.5$, we select the top-K most confident segments $\boldsymbol{X}_{1:K}^{(i)}$ of given record $\boldsymbol{x}^{(i)}$ for model fine-tuning process to achieve higher consistency among the augmented segments within a record and reduce the effects of noisy segments. Otherwise, we select the top-K segments of its inverted sequence. That is because a low average probabilistic prediction of all of its augmented segments indicate a poor performance of our model. Therefore, the prediction of our model is inaccurate and unreliable, and we need more "hard" segment samples for training. The selected α-segments $X_\alpha^{(i)}$ can be defined as follows:

$$\boldsymbol{X}_\alpha^{(i)} = \begin{cases} \boldsymbol{X}_{1:K}^{(i)}, & \text{if } \alpha_i > 0.5 \\ \boldsymbol{X}^{(i)} \setminus \boldsymbol{X}_{1:K}^{(i)}, & \text{otherwise} \end{cases} \tag{6}$$

Therefore, the task of AF detection using our K-margin-based RCR-net model can be expressed as optimizing the cross-entropy objective function:

$$L(X,Y) = \frac{1}{N|\boldsymbol{X}_\alpha^{(i)}|} \sum_{i=1}^{N} \sum_{t=1}^{|\boldsymbol{X}_\alpha^{(i)}|} \log P(Y = y_t^{(i)}|X = \boldsymbol{X}_{\alpha t}^{(i)}) \tag{7}$$

We further transform the label predictions $\hat{Y}_{1:K}^{(i)}$ of the top-K most confident segments for the given ECG record $\boldsymbol{x}^{(i)}$ under trained model ϕ to a matrix $\hat{\boldsymbol{Y}}_{1:K}^{(i)} \in \mathbb{R}^{K \times m}$ as follows:

$$\hat{\boldsymbol{Y}}_{k,j}^{(i)} = \begin{cases} 1, & \text{if } \hat{y}_k^{(i)} = c_j \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Thus, we propose a voting algorithm based on the top-K uncertainty variant measurement. To identify the most probable label of a given ECG record $\boldsymbol{x}^{(i)}$, we employ a K-margin-based diagnosis method to vote on the probability of rhythm class the ECG record $\boldsymbol{x}^{(i)}$ might belong to. The most probable label $\hat{y}^{(i)}$ of an ECG record $\boldsymbol{x}^{(i)}$ can be defined as follows:

$$\hat{y}^{(i)} = \underset{j}{\arg\max}(\sum_{1}^{K} \hat{\boldsymbol{Y}}_{k,j}^{(i)}) \tag{9}$$

Our K-margin-based RCR-net model can automatically focus on the most important segments associated with an ECG record and handle noise, as only a portion of the segments in each record participate in the learning process, by naturally exploiting expected consistency among the segments associated for each record. In our experiment, we demonstrate that our K-margin-based diagnosis can further improve the robustness and accuracy in the AF detection task.

4 Experiments and Results

4.1 Experimental Setup

Dataset

We carry out experiments on PhysioNet/Computing in Cardiology Challenge 2017 Dataset [Clifford *et al.*, 2017], which contains 8,528 single lead ECG records lasting from 9 s to just over 60 s sampling with 300 Hz. In this dataset, all non-AF abnormal rhythms are treated as a single class, namely Other rhythm. Thus these records are classified as 4 classes: 1) Normal sinus rhythm **N**, 2) Atrial Fibrillation **A**, 3) Other rhythm **O**, 4) Too noisy to classify **P**.

Evaluation Measurements

Commonly used multi-class metrics such as Precision, Recall and Hamming Loss are utilized to measure the performance by evaluating how close the predicting labels are to corresponding labels given by doctors:

- $Precision = \frac{1}{m} \sum_{c=1}^{m} \sum_{i \in \{i|y^{(i)}=c\}} \frac{\mathbb{1}(y^{(i)}=\hat{y}^{(i)})}{|\{i|y^{(i)}=c\}|}$.

- $Recall = \frac{1}{m} \sum_{c=1}^{m} \sum_{i \in \{i|y^{(i)}=c\}} \frac{\mathbb{1}(y^{(i)}=\hat{y}^{(i)})}{|\{i|\hat{y}^{(i)}=c\}|}$.

- $Hamming\ Loss = \frac{1}{N} \sum_{i=1}^{N} \frac{\mathbb{1}(y^{(i)} \neq \hat{y}^{(i)})}{m}$.

We also use the following F_1 evaluation measurements:

- F_1 *scores of each class*: Denoted as F_{1N} for normal sinus rhythm, F_{1A} for AF, F_{1O} for other rhythm, F_{1P} for noise. Detailed definitions can be found in [Clifford *et al.*, 2017].

- *Averages of* F_1 *scores*: $F_{1NAO} = \frac{F_{1N}+F_{1A}+F_{1O}}{3}$, $F_{1NAOP} = \frac{F_{1N}+F_{1A}+F_{1O}+F_{1P}}{4}$.

Implementation Details

In implementation, we randomly split 80% for model training, and evaluate on remaining 20% testing data. We evaluate the effects of various state-of-art DNN methods using the measurements given above by repeatedly running 20 times using 5-fold cross validation, and report the average results.

We set the parameter value of window_size and max stride threshold MS to be 6000 and 500 for skewness-driven dynamic data augmentation respectively. Then we set the parameter value of K and N_split to be 3 and 300 for training K-margin-based RCR-net using augmented data (see Section 3.2), and identifying the final predicting label of each ECG record by the trained model. Methods are implemented using Python 3.6.2 on TensorFlow version r1.4.

4.2 Results

Comparing with Other Methods

First of all, we compare our method with the following deep neural network methods:

- *RNN* is one of the most popular DNN architecture, which allows it to exhibit temporal dynamic behavior. It has been used as main classifier for detecting AF in [Warrick and Homsi, 2018].

- *CNN* use a variation of multilayer perceptrons to automatically extract features. Recently, [Sodmann *et al.*, 2018] constructed CNN architecture for AF detection without manual extracted features.
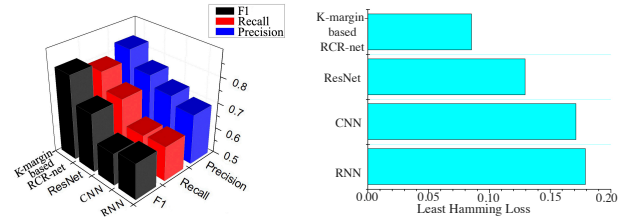

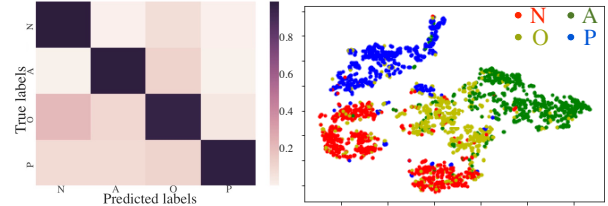Figure 5: (Left) The result of Precision, Recall, F1 scores. (Right) The result of hamming loss scores.

Figure 6: Confusion matrix and deep features visualization of RCR-net.

- *ResNet* allows training of very deep networks by constructing the network through modules called residual models, which can avoid the higher training error caused by naively adding the layers [He *et al.*, 2016]. In [Hannun *et al.*, 2019], they have achieved accuracies comparable to or higher than cardiologists for classifying 14 different types of arrhythmia by employing the ResNet as the main classifier.

Results are shown in Figure 5. The left part shows precision, recall and F_{1NAOP} results. We can see that the proposed K-margin-based RCR-net method with 0.8125 F_{1NAOP} score outperforms all state-of-the-art deep learning methods for AF detection task by 6.8%. The method of AlexNet and VGG are not effective, because it is difficult to construct very deep networks which will cause information loss and feature loss. The ResNet method works poorly because feature mapping simply aggregates all the automatic extracted local characters without considering any context information of overall long-term trend, which is constrained by the expressiveness of the model.

From the view of generalization ability of predictor (Hamming Loss), our K-margin-based RCR-net also performs better than others in terms of 38% - 50% fraction of labels that are incorrectly predicted, as shown in Figure 5 (right).

Besides, Table 1 also demonstrates our F_1 score and nine DNN methods employed in the Challenge 2017. We can see that automatically focusing on the most important segments (with a $6.8\% - 21.3\%$ higher F_{1NAOP} score when comparing it with other DNN methods) is an effective way to improve prediction performance (especially on handling noisy segments and identifying Too Noisy To Classify records).

Moreover, Figure 6 shows detailed performance of our method. The left part of Figure 6 shows a confusion matrix of the model predictions on the test set. Often mistakes made by the model are understandable. For example, many records of Other Rhythm are confused with Normal Sinus Rhythm

Method	F_{1N}	F_{1A}	F_{1O}	F_{1P}	F_{1NAO}	F_{1NAOP}
Ours	0.9061	**0.8712**	0.7166	**0.7561**	**0.8313**	**0.8125**
[Hong *et al.*, 2017]	**0.9117**	0.8128	**0.7505**	0.5671	0.8250	0.7605
[Zihlmann *et al.*, 2017]	0.9090	0.8221	0.7319	0.5676	0.8210	0.7577
[Xiong *et al.*, 2017]	0.9031	0.8203	0.7310	0.5251	0.8181	0.7449
[Schwab *et al.*, 2017]	0.9062	0.7385	0.7165	0.4751	0.7871	0.7091
[Andreotti *et al.*, 2017]	0.8923	0.7553	0.6715	0.4843	0.7730	0.7009
[Jiménez-Serrano *et al.*, 2017]	0.8990	0.7708	0.6944	0.4121	0.7881	0.6941
[Zihlmann *et al.*, 2017]	0.8884	0.7647	0.6686	0.4220	0.7739	0.6859
[Stepien and Grzegorczyk, 2017]	0.8973	0.7012	0.6420	0.4377	0.7468	0.6696
[Chandra *et al.*, 2017]	0.8600	0.7300	0.5600	N.A.	0.7167	N.A.

Table 1: Comparing our method with other deep learning methods in the PhysioNet/Computing in Cardiology Challenge 2017.
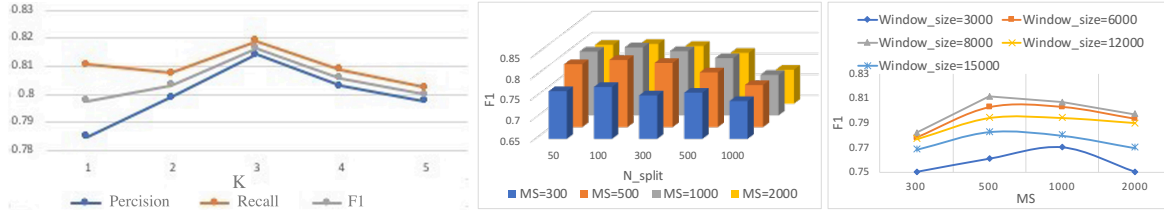
Figure 7: F_1 score results of K-margin-based RCR-net for different parameters.

which makes sense given that it is very difficult to distinguish between the normal sinus rhythm and some arrhythmias in the ECG records. Confusing Other Rhythm and Too Noisy To Classify also makes sense, as this can be subtle to detect especially when their ECG morphologies are similar or when noise is present. The right part of Figure 6 shows deep features learnt using the K-margin-based model on the test set. This shows whether the learned deep features effectively capture the latent relationships among the given four classes. Obviously, we can come to the conclusion that the learnt deep features are highly distinguished and representative.

Hyper-parameters Analysis

Now we further evaluating the performance of our method in different hyper-parameter settings.

The effects of threshold K are shown in the left part of Figure 7. We display the accuracy results of our K-margin-based RCR-net under different threshold K in the range of $[1, 3]$ and how the accuracy (Precision, Recall and F_1 score) raises dramatically. However, it drops when the K-margin threshold K is larger than 3. The reason is that the accuracy of our K-margin-based RCR-net model depends on both the number of segments engaged in the algorithm and the confidence of their labels. As K determines how many segments of an ECG record are engaged in the algorithm. Therefore, a little bit larger value of threshold K may result in higher accuracy. While the confidence of labelling accuracy decreases when K gets larger, as we select top-K segments from the most confident one to a less confident one. It means that higher K may also result in smaller confidence of labels engaged in the learning and voting phase, and therefore it would decrease the performance of our K-margin-based RCR-net model.

In addition, we evaluate the effects of three main parameters (windows_size, maximum stride threshold MS and

N_split). In the middle part of Figure 7, we display the accuracy results of our K-margin-based RCR-net model under different parameter N_split and MS in the range of $[50, 1000]$ and $[300, 2000]$ respectively. We can see that the model performs better by splitting ECG segments into smaller fragments other than larger fragments, as it could provide more fragments for recurrent neural network layer to capture rhythm-level trend features. However, the length of a fragment decreases with the increasing of the number of fragments which would cause the loss of some critical local heartbeat-level features.

Furthermore, as shown in the right part of Figure 7, when the model parameters of *window_size* and MS vary within the range of $[3000, 8000]$ and $[300, 2000]$ respectively, as smaller parameters increase the augmented training data, the K-margin-based RCR-net performs better. However, the performance decreases when using very small ones, as it can only provide a very narrow view of local beat-level features.

5 Conclusion

In this paper, we propose a K-margin-based RCR-net for AF detection. The experimental results demonstrate that the proposed method with 0.8125 F_{1NAOP} score outperforms all state-of-the-art deep learning methods for AF detection task by 6.8%. A possibly rewarding avenue of future research is to consider multi-modality data input and more fine-grained output categories to improve the model performance and apply it in a more practical situation.

References

[Andreotti *et al.*, 2017] Fernando Andreotti, Oliver Carr, and Marco AF Pimentel, et al. Comparing feature-based classifiers and convolutional neural networks to detect arrhyth-

mia from short segments of ecg. In *CinC*, pages 1–4. IEEE, 2017.

[Carrara *et al.*, 2015] Marta Carrara, Luca Carozzi, and Travis J Moss, et al. Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy. *Physiological measurement*, 36(9):1873, 2015.

[Chandra *et al.*, 2017] BS Chandra, Challa S Sastry, and Soumya Jana, et al. Atrial fibrillation detection using convolutional neural networks. In *CinC*, pages 1–4. IEEE, 2017.

[Clifford *et al.*, 2017] Gari D Clifford, Chengyu Liu, and Benjamin Moody, et al. Af classification from a short single lead ecg recording: the physionet/computing in cardiology challenge 2017. In *CinC*, pages 1–4. IEEE, 2017.

[Gradl *et al.*, 2012] Stefan Gradl, Patrick Kugler, and Clemens Lohmüller, et al. Real-time ecg monitoring and arrhythmia detection using android-based mobile devices. In *EMBC*, pages 2452–2455. IEEE, 2012.

[Hannun *et al.*, 2019] Awni Y Hannun, Pranav Rajpurkar, and Masoumeh Haghpanahi, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65, 2019.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, and Shaoqing Ren, et al. Deep residual learning for image recognition. In *CVPR*, pages 770–778. IEEE, 2016.

[Hong *et al.*, 2017] Shenda Hong, Meng Wu, and Yuxi Zhou, et al. Encase: An ensemble classifier for ecg classification using expert features and deep neural networks. In *CinC*, pages 1–4. IEEE, 2017.

[Hong *et al.*, 2019] Shenda Hong, Yuxi Zhou, and Meng Wu, et al. Combining deep neural networks and engineered features for cardiac arrhythmia detection from ECG recordings. *Physiological Measurement*, 40(5):054009, 2019.

[Jiménez-Serrano *et al.*, 2017] Santiago Jiménez-Serrano, Jaime Yagüe-Mayans, and Elena Simarro-Mondéjar, et al. Atrial fibrillation detection using feedforward neural networks and automatically extracted signal features. In *CinC*, pages 1–4. IEEE, 2017.

[Kamaleswaran *et al.*, 2018] Rishikesan Kamaleswaran, Ruhi Mahajan, and Oguz Akbilgic. A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using single lead electrocardiograms of variable length. *Physiological measurement*, 39(3):035006, 2018.

[Kass and Clancy, 2005] Robert E Kass and Colleen E Clancy. *Basis and treatment of cardiac arrhythmias*, volume 171. Springer Science & Business Media, 2005.

[Kiranyaz *et al.*, 2015] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675, 2015.

[Linker, 2016] David Linker. Accurate, automated detection of atrial fibrillation in ambulatory recordings. *CVET*, 7(2):182–189, 2016.

[Majumdar and Ward, 2017] Angshul Majumdar and Rabab Ward. Robust greedy deep dictionary learning for ecg arrhythmia classification. In *IJCNN*, pages 4400–4407. IEEE, 2017.

[Schwab *et al.*, 2017] Patrick Schwab, Gaetano C Scebba, and Jia Zhang, et al. Beat by beat: Classifying cardiac arrhythmias with recurrent neural networks. In *CinC*, pages 1–4. IEEE, 2017.

[Shah and Rubin, 2007] Atman P Shah and Stanley A Rubin. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. *Journal of electrocardiology*, 40(5):385–390, 2007.

[Silva *et al.*, 2011] Ikaro Silva, George B Moody, and Leo Celi. Improving the quality of ecgs collected using mobile phones: The physionet/computing in cardiology challenge 2011. In *CinC*, pages 273–276. IEEE, 2011.

[Sodmann *et al.*, 2018] Philipp Sodmann, Marcus Vollmer, and Neetika Nath, et al. A convolutional neural network for ecg annotation as the basis for classification of cardiac rhythms. *Physiological measurement*, 39(10):104005, 2018.

[Stepien and Grzegorczyk, 2017] Katarzyna Stepien and Iga Grzegorczyk. Classification of ecg recordings with neural networks based on specific morphological features and regularity of the signal. In *CinC*, pages 1–4. IEEE, 2017.

[Tateno and Glass, 2001] K Tateno and L Glass. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of rr and δrr intervals. *MBEC*, 39(6):664–671, 2001.

[Warrick and Homsi, 2018] Philip A Warrick and Masun Nabhan Homsi. Ensembling convolutional and long short-term memory networks for electrocardiogram arrhythmia detection. *Physiological measurement*, 39(11):114002, 2018.

[Wu *et al.*, 2016] Zhiyong Wu, Xiangqian Ding, and Guangrui Zhang. A novel method for classification of ecg arrhythmias using deep belief networks. *IJCIA*, 15(04):1650021, 2016.

[Xiong *et al.*, 2017] Zhaohan Xiong, Martin K Stiles, and Jichao Zhao. Robust ecg signal classification for detection of atrial fibrillation using a novel neural network. In *CinC*, pages 1–4. IEEE, 2017.

[Ye *et al.*, 2010] Can Ye, Miguel Tavares Coimbra, and BVK Vijaya Kumar. Arrhythmia detection and classification using morphological and dynamic features of ecg signals. In *EMBC*, pages 1918–1921. IEEE, Aug 2010.

[Zihlmann *et al.*, 2017] Martin Zihlmann, Dmytro Perekrestenko, and Michael Tschannen. Convolutional recurrent neural networks for electrocardiogram classification. In *CinC*, pages 1–4. IEEE, 2017.