Hybrid Arrhythmia Detection on Varying-Dimensional Electrocardiography: Combining Deep Neural Networks and Clinical Rules

Hao Wen[1], Jingsu Kang[2]

[1]School of Mathematical Sciences, Beihang University, Beijing, China
[2]Tianjin Medical University, Tianjin, China

Abstract

Aim: This study (from Revenger team) aims to develop effective approaches for the detection of cardiac arrhythmias from varying-dimensional electrocardiography (ECG) in the PhysioNet/Computing in Cardiology Challenge 2021, taking advantage of both deep neural networks (DNNs) and insights from clinical diagnostic criteria.

Methods: 26 classes (equivalent classes are counted one) of ECGs are divided into two categories. Detectors are manually designed for classes in the category with clear clinical rules. The rest classes with subtle morphological and spectral characteristics are classified by DNNs. To make the networks capable of capturing features of different scopes, we use multi-branch convolutional neural networks (CNNs), each with different receptive fields via dilated convolutions. Considering ECGs' varying dimensionality, convolutions are grouped with group number equaling the number of leads. Outputs from DNNs and from manual detectors are merged to give final predictions.

Results: Although we did not officially rank (the code failed to complete on the 12-lead test set), we received test scores of 0.33, 0.35, 0.33, 0.33, and 0.33 on the 2-lead, 3-lead, 4-lead and 6-lead test sets respectively.

Conclusion: The proposed hybrid method is effective for establishing auxiliary diagnosis systems, and the reduced-lead ECGs are sufficient for such systems.

1. Introduction

The electrocardiogram (ECG), a physiological signal reflecting the electrical activities of cardiac muscles, is a major tool for screening and early intervention of cardiac diseases which is the leading cause of death worldwide [1]. However, data amount generated by ECG apparatus are typically large, especially the 24-hour Holter monitors. Hence accurate automated auxiliary diagnosis systems of cardiac electrical abnormalities from ECGs is crucial.

The PhysioNet/Computing in Cardiology Challenge 2021 (CinC2021) focused on such mission, or more precisely automated, open-source approaches for classifying cardiac abnormalities from reduced-lead (reduced from the standard 12-lead) ECGs [2–4]. In this paper, our effort of tackling this problem, which used DNNs combined with clinical rules, will be described.

2. Methods

2.1. Partitioning and Selection of Data

Coarsely, the scored ECG arrhythmias of the challenge can be divided into 2 categories. Most classes (21 classes, with equivalent classes counted as one) are characterized by subtle morphological and spectral changes. This majority of ECG arrhythmias are detected (classified) using DNNs, which are described in Section 2.3. The rest 5 ("Brady", "LAD", "RAD", "LQRSV", "PR") have clear and easy-to-describe clinical diagnostic criteria. For these ECG arrhythmias, manually designed detectors from clinical rules are included as a part of our solution. These detectors are described in more details in Section 2.5.

Although there are totally 132 classes (abnormalities) available in the challenge database, only the scored ones are included for the development of our challenge approach. ECG records with no scored classes are discarded. We also exclude the StPetersburg subset (the INCART dataset [5]) from training the models for several reasons. Most importantly, these records are 30 minutes long with only at most 2 classes of scored abnormalities, which is too coarse. Second, each of the 9 scored classes constitutes less than 0.2% of the total challenge database, which is almost neglectable.

2.2. Preprocess and Data Augmentation

To make training and inference data in better consistency, data are filtered using a Butterworth filter of order 5 and passband 0.5 Hz - 60 Hz, after which baseline wander and high frequency noises are removed. The high cutoff frequency is slightly higher than usual due to the fact

that the distinguishing characteristics of the pacing rhythm ("PR") are vertical spikes of very short duration.

For training DNNs, the ECGs are resampled to 500 Hz, cropped or zero-padded to ensure 10-second length (5000 sample points) to utilise mini-batch (parallel) training. In our approaches, if detection of the class "LQRSV" (low qrs voltages) is not included in the task of CNN models, the input ECGs are further normalized to have zero mean and unit variance, since this abnormality is directly related to absolute values in voltages. To alleviate overfitting, random masking [6] with zero values of ECG segments of length at most 1.0s is adopted. No more augmentations, like random flip, are done, since they might completely change the interpretation of the ECGs, as will be seen in Section 2.5. To suppress overconfidence which could help improving generalization capabilities of models, the technique of label-smoothing regularization (LSR) [7] is used. Let $\mathbf{y}$ be an one-hot label vector (the "hard" label), then LSR generates "soft" label vector via Equation (1)

$$\mathbf{y}' = (\mathbf{1} - \varepsilon)\mathbf{y} + \frac{\mathbf{1}}{\mathbf{K}}\varepsilon\mathbf{e}, \tag{1}$$

where K is the number of classes, $\mathbf{e}$ the K-dimensional vector with all entries equaling one, and $\varepsilon \in [0, 1]$ is a weight factor. In our approach, we take $\varepsilon = 0.1$.

2.3. Neural Network Architectures

Convolutional Recurrent Neural Networks (CRNNs). Inspired by previous work [6, 8], we build our approach on top of a CRNN framework[1]. The philosophy is as follows.

CNNs consist of space translation equivariant convolution operators, which usually interleave with non-linear activations and downsampling operators capturing and fusing hierarchical local features. In our CRNN framework, CNNs serve as feature extractors (encoders) from raw input ECGs. In order to better modeling long-range dependency, optional (self-)attention modules (SENet [9], etc.) can follow or integrated in building block convolutions in the CNNs. An optional RNN can be added as well to make use of sequential information of the ECGs. Feature maps thus obtained are fed into multilayer perceptrons (MLPs) for ECG downstream tasks, including classification, sequence labeling (e.g. QRS complex detection [8]), etc.

Multi-branch CNNs. The most significant structures of ECGs are the P, Q, R, S, T waves and their rhythms which for example can be reflected by the sequence of wave intervals (RR intervals, PP intervals, QT intervals, etc.). Broadly speaking, these waves and intervals broadly have their "general" spectral characteristics which originate from the mechanism of the human heart's electric activities. Hence the receptive fields of the CNNs is crucial for ECGs, or more widely for physiological signal processing. Previous work [8] explicitly models this via multi-branch CNNs where each branch uses different dilation factors. Therefore, we mainly experimented with multi-branch CNNs in our approach.

[1] indeed an ECG deep learning framework more broadly, available at https://github.com/DeepPSP/torch_ecg

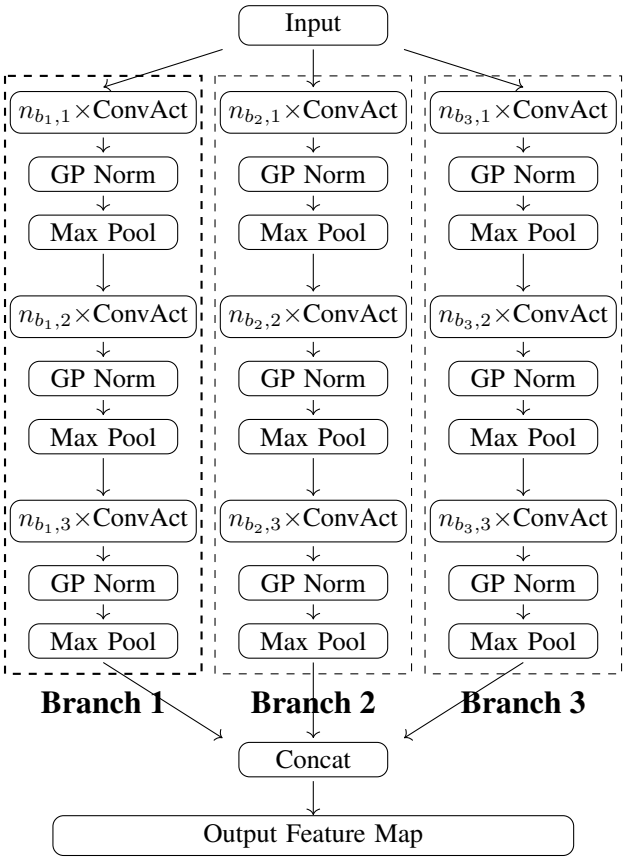

Figure 1. A typical 3-branch CNN. Abbreviations: "ConvAct" for grouped convolution layer followed by ReLU activation layer, "GP Norm" for group normalization layer, "Max Pool" for max pooling layer of kernel size 2. We set $(n_{b_i,1}, n_{b_i,2}, n_{b_i,3}) = (1, 2, 3)$, $i = 1, 2, 3$, in our challenge entry approaches. The six convolution layers of each branch have kernel sizes $(11, 7, 7, 5, 5, 5)$, and dilation factors of $(1, 1, 1, 1, 1, 1)$, $(2, 2, 4, 8, 8, 8)$, and $(4, 4, 8, 16, 32, 64)$ from branch 1 to branch 3. This Architecture can shrink or expand horizontally by removing or adding branches, and shrink or expand vertically by removing or appending convolutions.

"Lead-wise" CNNs. The challenge [4] emphasises the utility of reduced-lead ECGs, hence in our approach we designed a "lead-wise" manner for the CNNs via grouped convolutions with number of groups divisible by the number of leads of the input ECGs. For example, 12 groups for the standard 12-lead ECGs. In this "lead-wise" settings, normalization layers are group normalizations [10] as well.

# leads	12	6	4	3	2
$n_{b_i,1}\times$Conv	192	144	96	96	64
$n_{b_i,2}\times$Conv	384	288	192	192	128
$n_{b_i,3}\times$Conv	768	576	384	384	256

Table 1. Number of filters for the convolutions in the CNN described in Figure 1.

$$\text{CNN} \xrightarrow[\text{feature maps}]{\text{separated}} \text{Attention} \xrightarrow[\text{feature maps}]{\text{fused}} \text{MLP Head}$$

Figure 2. The whole network architecture. CNN is described in Figure 1. The attention module used in our approach is SENet with reduction ratio 8. Then adaptive max pooling is applied to the fused feature maps to reduce the number of channels to one. Finally, MLP consisting of one linear layer gives the predictions which are tensors of probabilities for each of the classes.

In this way, CNNs extract features for each lead separately in parallel. Features from different leads are not fused until forwarded out from the CNNs. This provides the possibility to reuse parameters from the models trained on the standard 12-lead ECGs for reduced-lead ECGs, in which case one only needs to "fine-tune" the attention modules and the MLPs. This can play the role of general-purposed "backends" as in computer vision. Another advantage is that "lead-wise" CNNs are much smaller in the number of model parameters, with only slight drop of performance.

The major CNN architecture in our challenge entry approaches is plotted in Figure 1 with number-of-leads-independent hyperparameters included therein. These hyperparameters are inherited from [8] directly. The number of filters are listed in Table 1.[2] The whole network is gathered in Figure 2.

2.4. Training Setups

Since the challenge data is highly unbalanced, having a long tail distribution, we thus use weighted binary cross entropy as the loss function. The weights are inversely proportional to the number of records of the classes. 20% of the training data is left out for validation and model selection. We set batch sizes 32 or 64 depending on the model sizes, and set the maximum number of training epochs to be 30 with early stopping. Model parameters are optimized using the AMSGrad variant of the AdamW optimizer [11] with learning rate 0.001. To make binary predictions from probabilities, a threshold 0.5 is used. If none exceeds 0.5, then the class with the highest probability and classes with close enough (within a bias of 0.03) probabilities are chosen as the binary output.

[2]Mechanism of reuse of parameters for reduced-lead ECGs has not been established by the end of the challenge, hence we set this hyperparameter for each of the 5 lead sets.

2.5. Clinical Rule Based Detectors

Clinical rules based detectors are designed for the 5 ECG abnormalities listed in Section 2.1. From the authors' experiences of previous challenges and production systems, post-processing using clinical rules is an excellent supplement to machine learning models. Details are:
1. "Brady": average heart rate $\leq$ 60 BPM (beat per minute) or equivalently average RR-intervals $\geq$ 1 second.
2. "LAD" and "RAD": positivity checking of QRS complexes of leads I, aVF ("2-lead" method) as in [12].
3. "LQRSV": peak-to-peak amplitudes of more than 80% of the QRS complexes are $\leq$ 0.5 mV in the limb leads (I, II, III, aVR, aVL, aVF), or $\leq$ 1 mV in the precordial leads (V1-V6). If R peak detection fails, amplitude check will be done within sliding windows of length 0.12 second.
4. "PR": raw ECGs are high-pass filtered with cutoff frequency 47 Hz, and spike (peak) detection with prominence threshold of 0.3 follows.

Detection of the first 4 abnormalities relies heavily on R peak detection, for which we use the function "xqrs_detect" from the WFDB package [2, 13] for simplicity. This function however is far from optimal, causing miss-classifications. The 5-dimensional outputs and the 21-dimensional outputs from DNNs are naively merged to produce the final predictions.

3. Results

Our challenge entries mainly uses two configurations, namely the proposed hybrid method, and the pure DNN approach. Best scores of challenge entry submissions and offline experiments are gathered in Table 2.

We carried out offline expriments using CNNs without the "lead-wise" setting, but no successful entry submission was made, perhaps due to the increase of model size that exceed the computation capacity. These ablation studies results hence are not reported in this paper.

4. Discussion and Conclusions

The hybrid method of DNNs and clinical rules provides effective approaches for automated auxiliary multi-lead ECG diagnosis systems. It can be inferred from the results that reduced-lead ECGs, even 2-lead ECGs in the extreme case, provide sufficient information for making reliable auxiliary diagnoses, with performances (challenge metric) only slightly dropped by at most 0.03, compared to the standard 12-lead ECGs on the validation set.

There are limitations and left for future work. First, the multi-branch CNNs for feature extraction are far from op-

Leads	Training	Validation	Test	Ranking
12	0.62	0.51	NA	NA
6	0.59	0.47	0.33	NA
4	0.60	0.47	0.35	NA
3	0.61	0.48	0.33	NA
2	0.59	0.48	0.33	NA
12-cr	0.64	0.51	NA	NA
6-cr	0.61	0.49	NA	NA
4-cr	0.61	0.44	NA	NA
3-cr	0.61	0.46	NA	NA
2-cr	0.59	0.43	NA	NA

Table 2. Challenge scores (top 5 rows) for the final entry. The final entry used the "no clinical rule" pure DNN approach. It failed on the 12-lead test set (more exactly the 12-lead "UMich test" set). The auxiliary bottom 5 rows (with "-cr" suffix) describes performances of our hybrid approach mixing DNN and clinical rules. Scores in the "Training" column are typical scores on the 20% left-out train-validation data, as described in Section 2.4.

timal, compared to approach of other challenge teams. Its structures and hyperparameters both have to be optimized. A thorough search for more effective architectures should and is undertaken by the authors in the ECG deep learning framework mentioned in Section 2.3. Second, it is observed in Table 2 that performances of the hybrid entries on the reduced 4-lead, 3-lead and 2-lead ECGs dropped slightly larger than pure DNN entries. The hyperparameters of clinical rule based detectors are set empirically, which should be optimized via grid searches. Label heterogeneity and insufficiency across datasets should also be noted. We observed labels that violates clinical criteria, for example some "LAD" records violates the "3-lead" method which is more exact than the "2-lead" method mentioned in Section 2.5.

Most importantly, the "lead-wise" CNNs provides flexible light weight solutions to reduced-lead ECGs. The mechanism of parameters reuse is to be further established.

Acknowledgments

The authors would like to thank professor Deren Han from the School of Mathematical Sciences, Beihang University and professor Wenjian Yu from the Department of Computer Science and Technology, BNRist, Tsinghua University for generously providing GPU servers to help accomplish this work.

References

[1] Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart Disease and Stroke Statistics – 2021 Update: a Report from the American Heart Association. Circulation 2021;143(8):e254–e743.

[2] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215–e220.

[3] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement Nov 2020;41. Doi: 10.1088/1361-6579/abc960.

[4] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48:1–4.

[5] Tihonenko V, Khaustov A, Ivanov S, Rivin A, Yakushenko E. St Petersburg INCART 12-lead Arrhythmia Database. PhysioBank PhysioToolkit and PhysioNet 2008;Doi: 10.13026/C2V88N.

[6] Yao Q, Wang R, Fan X, Liu J, Li Y. Multi-class Arrhythmia Detection from 12-lead Varied-length ECG Using Attention-based Time-Incremental Convolutional Neural Network. Information Fusion 2020;53:174–182.

[7] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016; 2818–2826.

[8] Cai W, Hu D. QRS Complex Detection Using Novel Deep Learning Neural Networks. IEEE Access 2020;8:97082–97089.

[9] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 2020;42(8):2011–2023.

[10] Wu Y, He K. Group Normalization. International Journal of Computer Vision 7 2019;128(3):742–755. Doi: 10.1007/s11263-019-01198-w.

[11] Reddi SJ, Kale S, Kumar S. On the Convergence of Adam and Beyond. In 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada. OpenReview.net, 2018; .

[12] Kashou AH, Basit H, Chhabra L. Electrical Right and Left Axis Deviation. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, Jan 2021. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470532/.

[13] Xie C, McCullum L, Johnson A, Pollard T, Gow B, Moody B. Waveform Database Software Package (WFDB) for Python, 2021. Doi: 10.13026/G35G-C061.

Address for correspondence:

Hao Wen
E301, Beihang University, Changping District, Beijing, China
wenh06@buaa.edu.cn, wenh06@gmail.com