Automatic Diagnosis of Cardiac Disease from Twelve-Lead and Reduced-Lead ECGs Using Multilabel Classification

Prathic Sundararajan, Kevin Moses, Cristhian Potes, Saman Parvaneh

Edwards Lifesciences, Irvine, USA

Abstract

ECG is an essential tool for the clinical diagnosis of cardiac electrical abnormalities. As part of the PhysioNet/Computing in Cardiology Challenge 2021, eight and two folds from the 10-folds iterative splitting of public training data set were used as in-house training and internal validation sets. We used extracted features from RandOm Convolutional KErnel Transforms (ROCKETs) with a multilabel classification using XGBoost to predict cardiac abnormalities. Our team, LINC, developed an approach with minimal pre-processing (e.g., resampling data to 500Hz) and with no QRS detection or deep neural network design, which led to promising performance on the internal validation set. We didn't receive the official scores for the validation and test sets, because our entry failed during training in the official phase as we submitted an incomplete entry. Our classifiers received scores of 0.504, 0.466, 0.459, 0.458, and 0.438 for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead versions on the internal validation set with the challenge evaluation metric (10 seconds ECG).

1. Introduction

Cardiovascular disease is the leading cause of death worldwide resulting in significant health and economic burden in the United States and globally [1]. The electrocardiogram (ECG) is an essential non-invasive tool for the clinical diagnosis of cardiac diseases [2]. The 2021 PhysioNet/Computing in Cardiology Challenge aims to develop automated, open-source algorithms to classify cardiac abnormalities from twelve-lead and reduced lead ECGs [3-5]. Traditional algorithms for cardiac disease classification require ECG signal processing (e.g., filtering and QRS detection), feature extraction using domain knowledge, and classifier development [6].

Furthermore, deep learning models have produced promising classification performance for cardiac abnormalities without the need for handcrafted features provided by domain knowledge [7-9]. However, developing a best-performing deep learning model requires a trial-and-error process to find optimum network architecture and parameters. A classification using features learned by a RandOm Convolutional KErnel Transforms (ROCKETs) has shown high performance in different time series classifications [10]. Unlike deep learning model development, there is no need to identify and optimize a network architecture when using random convolution kernels with classifiers. Our proposed method for the challenge used ROCKET features with a multilabel classification using XGBoost for cardiac disease classification.

2. Material and Method

2.1 Data

Provided training data for the challenge is 12-leads ECG from six datasets: CPSC and CPSC-Extra database [11], INCART database [12], the PTB database and the PTB-XL database [13, 14], the Chapman-Shaoxing Database and the Ningbo Database [15, 16], the Georgia 12-lead ECG Challenge (G12EC) Database, and Augmented Undisclosed Database [3-5]. Details about the challenge dataset can be found in [3, 5]. Each of the ECG recordings has one or more labels that describe either a cardiac abnormality or a normal sinus rhythm. The challenge organizers mapped the labels to SNOMED-CT codes and kept the following four abnormality pairs as equivalent in terms of classification: CLBB and LBB, CRBB and RBB, PAC and SVPB, PVC and VPB. Further detailed information about the mapping of the labels can be found in [3, 5].

Iterative stratification using scored labels is used to split the training data into ten folds [3-5] while keeping prevalence of scored labels across different folds. Eight and two folds were selected randomly as in-house training and internal validation sets, respectively. To evaluate robustness of results, model training and evaluation on training and internal validation set repeated two times.

2.2 Method

A block diagram of our proposed algorithm is shown in Figure 1.

 Page 1 ISSN: 2325-887X DOI: 10.22489/CinC.2021.059

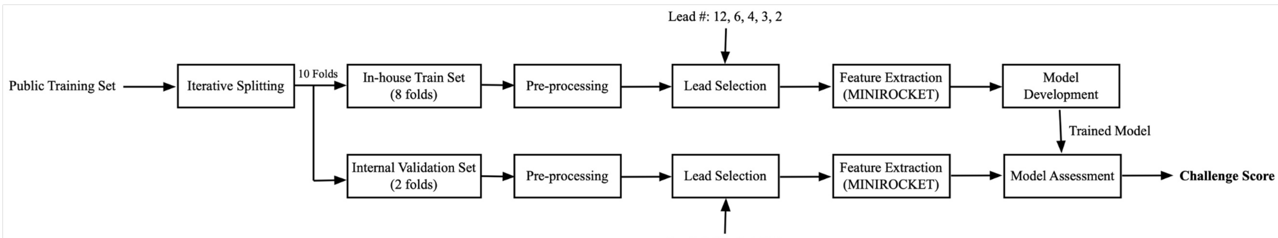

Figure 1. Block diagram of the proposed algorithm.

2.2.1 Pre-processing

Using provided gain and baseline in the ECG header, ECG amplitude was corrected to a physical unit. Considering the difference in sampling frequency across different datasets, all ECGs were resampled to 500Hz whenever necessary. We extracted 10 seconds from each recording and applied zero padding if ECG was shorter than 10 seconds.

2.2.2 Feature Extraction

Transformation-based classification methods transform the time series (e.g., ECG) using a transformation kernel and then use a classifier [17]. In our proposed algorithm, ROCKET randomly initializes many convolutional kernels (10,000) and uses them to transform the data. Each kernel is defined by $[l, w_i(i = 1, \ldots., l), b, d, p]$ where l is the length of the kernel and w_i, b, and d are kernel weights, bias, and dilation, respectively. Also, p is a boolean determining if padding is used or not.

In ROCKET, two features from each convolution between input signal and kernel are extracted for feature extraction (20,000 features) followed by a classifier (section 2.2.3). These features are:

1- Maximum of the convolution operation:

$$MAX = (s[n] * k[n]) \quad (1)$$

2- Proportion of positive values (PPV)

$$PPV = \frac{1}{N}\sum_{m=0}^{N-1}[(s[n] * k[n])_m + b > 0] \quad (2)$$

Where $*$, $s[n]$, $k[n]$, and b are convolution operation, input signal, kernel, and the bias scalar, respectively. $[Q]$ will be 1 when Q is true and 0 otherwise. The PPV provides the proportion of the input signal correlated with the kernel.

In this paper, MINIROCKET, which is a reformulated version of ROCKET, is used. It is shown that MINIROCKET with using only PPV is faster than ROCKET while maintaining the same accuracy [18]. Of note, MINIROCKET map input signal to 10,000 features that could be used with a classifier. An extension of MINIROCKET for using it with multivariate datasets (e.g., multi-lead ECG) implemented in the sktime library was used in this paper [19].

2.2.3 Model Development

We used SNOMED CT codes used in the challenge score, and all unscored classes were excluded from model development. Six percent of the total training dataset, consisting of 6,510 ECG records, had no cardiac arrythmia label since we were only classifying scored labels as specified by the challenge. About 40% of the training dataset, consisting of 32,200 ECG records, had more than one scored diagnosis. Therefore, multilabel classification was selected for ECG classification. The label powerset approach [20] was used to transform a multilabel problem into a multi-class problem—one multi-class classifier XGBoost [21] trained on all unique label combinations found in the training data for scored classes.

2.3 Model Evaluation

Thirty diagnoses were used in the challenge evaluation metric, and you can refer to [4, 5] for more details. An in-house training set (eight folds) is used for model training, and an internal validation set (two folds) is used for model evaluation before submissions. In the model development, 5 and 10 seconds of ECG were used as an input to the model.

3. Results

Our entry was not scored because of error in our code during the official phase as we failed to provide the "weights.csv" file in our repository. As a result, we didn't receive the official scores for the validation and test sets. Results on internal validation set for the model training on 5 and 10 seconds of data are reported in Table 1.

As reported in Table 2, for the model training on 10 seconds of data, challenge scores on the internal validation set for 12-leads, 6-leads, 4-leads, 3-leads, and 2-leads were 0.504, 0.466, 0.459, 0.458, and 0.438, respectively (Team: LINC). Additionally, F1 score for classification of each cardiac abnormality are calculated to identify where the developed models perform well and when they fail. Ranking by F1 score, Table 3 shows the twelve lead and reduced-lead model's best and worst performing diagnoses, respectively.

Leads	5 Seconds	10 Seconds
12	0.499	**0.504**
6	**0.473**	0.466
4	**0.459**	**0.459**
3	**0.460**	0.458
2	0.406	**0.438**

Table 1. Challenge scores for internal validation set with 5- and 10-seconds ECG as an input.

Leads	Internal validation set	Validation	Test	Ranking
12	0.504	NS	NS	NS
6	0.466	NS	NS	NS
4	0.459	NS	NS	NS
3	0.458	NS	NS	NS
2	0.438	NS	NS	NS

Table 2. Challenge scores for our best entry (team *LINC*) using internal validation set on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set as well as the ranking on the hidden test set. (NS: not officially scored because of errors in our code during the official phase).

4. Discussion and Conclusions

We proposed a multilabel XGBoost for classifying cardiac abnormalities from twelve-lead and reduced-lead ECGs using RandOm Convolutional KErnel Transform (ROCKET) with 10,000 kernels for feature extraction. Promising results on the internal validation set on the public training set indicate the power of ROCKET for feature extraction. Creating many kernels in ROCKET allows it to match patterns with complex shapes and frequencies within ECG data [21].

Except for the number of kernels in MINIROCKET, there are no other parameters to select or optimize [10, 19]. Therefore, cardiac abnormality classification in ECG can be done with no design or domain knowledge, irrespective of traditional and deep learning methods. As a limitation, features generated by MINIROCKET are not easy to be interpreted and future work is required to make ROCKET interpretable.

In the original ROCKET paper, PPV was a more important feature than MAX [10], and therefore we used MINIROCKET in this paper [18]. However, a comparison between ROCKET that uses both *MAX* and *PPV* and MINIROCKET for cardiac abnormality prediction needs to be evaluated in future studies.

As Table 1 shows, an increase in the length of ECG recording data used for feature extraction by MINIROCKET led to an increase in score for the twelve-lead model and two lead model. However, a trend in the opposite direction is observed for the six lead and three lead model. Further work should be done focusing on the impact of the recording length for feature extraction using MINIROCKET to better understand its effects on efficiency and performance of the model.

Leads	Four Highest F1 Diagnosis (F1 Score)	Four Lowest F1 Diagnosis (F1 Score)
12	SB (0.957) STach (0.915) NSR (0.899) PR (0.851)	SVPB & PAC (0.083) Brady (0.081) PRWP (0.029) LPR (0.001)
6	SB (0.946) STach (0.906) NSR (0.891) PR (0.850)	NSIVCB (0.060) IRBBB (0.025) LPR (0.020) PRWP (0.001)
4	SB (0.948) STach (0.901) NSR (0.884) PR (0.816)	NSIVCB (0.065) LPR (0.040) Brady (0.001) PRWP (0.001)
3	SB (0.941) STach (0.897) NSR (0.872) PR (0.809)	NSIVCB (0.044) Brady (0.001) LPR (0.001) PRWP (0.001)
2	SB (0.941) STach (0.897) NSR (0.872) PR (0.809)	NSIVCB (0.033) LPR (0.020) Brady (0.001) PRWP (0.001)

Table 3. Four abnormalities with lowest and highest F1 score in the classification task for the different lead combinations are listed with the F1 score shown inside the parenthesis. (Sinus Bradycardia: SB, Sinus Tachycardia: STach, Sinus Rhythm: NSR, Pacing Rhythm: PR, Supraventricular Premature Beats: SVPB, Premature Atrial Contractions: PAC, Bradycardia: Brady, Poor R Wave Progress: PRWP, Prolonger PR Interval: LPR, Nonspecific Intraventricular Conduction Disorder: NSIVCB, Incomplete Right Bundle Branch Block: IRBBB)

Additionally, it is observed on Table 3, that developed models perform well on abnormalities impacting ECG morphology. Comparatively, models performed poorly on ECG diagnoses that require interval measurements such as RR intervals (heart rate) and PR intervals. This might be due to MINIROCKET's focus on morphological pattern discovery using convolutional kernels. In the future study, we would like to evaluate the impact of combining ROCKET/MINIROCKET features with handcrafted features (e.g., heart rate variability and entropy). Also, the impact of replacing XGBoost with a linear classifier used in the original ROCKET paper (e.g., ridge regression classifier) on challenge score needs to be addressed in the future.

References

[1] A. Alonso, F. H. J. Aparicio, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway *et al.*, "Heart disease and stroke statistics—2021 update," *Circulation,* vol. 2021, no. 143, pp. e00-e00, 2021.

[2] P. Kligfield, "The centennial of the einthoven electrocardiogram," *Journal of Electrocardiology,* vol. 35, no. 4, pp. 123-129, 2002.

[3] E. A. P. Alday, A. Gu, A. J. Shah, C. Robichaux, A.-K. I. Wong *et al.*, "Classification of 12-lead ECGs: the Physionet/Computing in Cardiology Challenge 2020," *Physiological Measurement,* vol. 41, no. 12, p. 124003, 2020.

[4] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation,* vol. 101, no. 23, p. e215, 2000.

[5] M. A. Reyna, Sadr, N., Alday, E.A.P., Gu, A., Shah, A.J., Robichaux, C., Rad, A.B., Elola, A., Seyedi, S., Ansari, S., Ghanbari, H., Li, Q., Sharma, A., and Clofford, G.D., "Will two do? varying dimensions in electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021," in *Computing in Cardiology 2021*, 2021, vol. 48, pp. 1-4.

[6] S. Parvaneh and J. Rubin, "Electrocardiogram Monitoring and Interpretation: From Traditional Machine Learning to Deep Learning, and Their Combination," presented at the Computing in Cardiology 2018, 2018.

[7] S. Parvaneh, J. Rubin, S. Babaeizadeh, and M. Xu-Wilson, "Cardiac arrhythmia detection using deep learning: a review," *Journal of Electrocardiology,* vol. 57, pp. S70-S74, 2019.

[8] S. Parvaneh, J. Rubin, A. Rahman, B. Conroy, and S. Babaeizadeh, "Analyzing single-lead short ECG recordings using dense convolutional neural networks and feature-based post-processing to detect atrial fibrillation," *Physiological Measurement,* vol. 39, no. 8, p. 084003, 2018.

[9] J. Rubin, S. Parvaneh, A. Rahman, B. Conroy, and S. Babaeizadeh, "Densely connected convolutional networks for detection of atrial fibrillation from short single-lead ECG recordings," *Journal of Electrocardiology,* 2018.

[10] A. Dempster, F. Petitjean, and G. I. Webb, "ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining and Knowledge Discovery,* vol. 34, no. 5, pp. 1454-1495, 2020.

[11] C. Lundy and J. M. O'Toole, "Random convolution kernels with multi-scale decomposition for preterm EEG inter-burst detection," *arXiv preprint arXiv:2108.02039,* 2021.

[12] V. Tihonenko, A. Khaustov, S. Ivanov, A. Rivin, and E. Yakushenko, "St Petersburg INCART 12-lead arrhythmia database," *PhysioBank, PhysioToolkit, and PhysioNet,* 2008.

[13] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet," *Biomedizinische Technik,* no. 40(S1), pp. 317-318, 1995 1995.

[14] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data,* vol. 7, no. 1, pp. 1-15, 2020.

[15] J. Zheng, H. Chu, D. Struppa, J. Zhang, M. Yacoub *et al.*, "Optimal multi-stage arrhythmia classification approach," *Scientific Reports,* vol. 10, no. 1, pp. 1-17, 2020.

[16] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific Data,* vol. 7, no. 1, pp. 1-8, 2020.

[17] A. Shifaz, C. Pelletier, F. Petitjean, and G. I. Webb, "Elastic similarity measures for multivariate time series classification," *arXiv preprint arXiv:2102.10231,* 2021.

[18] A. Dempster, D. F. Schmidt, and G. I. Webb, "MiniRocket: A very fast (almost) deterministic transform for time series classification," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 248-257.

[19] M. Löning, A. Bagnall, S. Ganesh, V. Kazakov, J. Lines, and F. J. Király, "Sktime: a unified interface for machine learning with time series," *arXiv preprint arXiv:1909.07872,* 2019.

[20] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition,* vol. 37, no. 9, pp. 1757-1771, 2004.

[21] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd Acm Sgkdd International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

Address for correspondence:
Saman Parvaneh
1 Edwards Way, Irvine, CA, USA
parvaneh@ieee.org