

Recognition of Child Congenital Heart Disease Using Cardiac Cycle Segment of Electrocardiogram

Yunmei DU^a, Canhui HUANG^a, Shuai HUANG^b and Huiying LIANG^{b,1}

^a*School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou, China*

^b*Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Guangzhou, China*

Abstract. The results of previous studies showed that ECG could detect CHD in children with a detection rate of 76.43%. Although this result is better than the traditional CHD screening method, the sensitivity still needs to be improved if it is to be popularized clinically. Based on the previous ECG recording data, this study selects the more representative cardiac cycle segments to identify CHD, in order to achieve better screening effect. Firstly, better cardiac cycle segment data were extracted from ECG records of each patient. The final data set contains 72626 patients and each patient has a 9-lead ECG segment with duration of about one second. Then we trained a RoR network to identify the underlying patients with CHD using 62626 samples in the dataset. When tested on an independent set of 10000 patients, the network model yielded values for the sensitivity, specificity, and accuracy of 0.93, 86.3%, 85.7%, and 85.7% respectively. It can be seen that extracting more effective cardiac cycle fragments can significantly improve the sensitivity of CHD screening on the basis of ensuring better specificity, so as to find more potential patients with congenital heart disease.

Keywords. Electrocardiography (ECG), congenital heart disease (CHD), deep Learning, cardiac cycle segment

1. Introduction

Early screening of congenital heart disease is particularly important [1]. Traditional screening methods cannot fully meet the needs of large-scale screening because of cost, medical resources, detection rate and timeliness [2-5]. ECG is an early routine examination item, which is easy to obtain. Our previous study has verified the feasibility of using ECG to identify CHD patients, which provides a new idea for CHD screening [6].

This study evaluated whether the manually intercepted ECG cardiac cycle segment can have better performance than the whole ECG data, and whether it can improve the performance of the screening model.

¹ Corresponding Author, Huiying LIANG, Guangdong Provincial People's Hospital, Guangzhou, China; Email: lianghuiying@hotmail.com.

2. Data Preprocessing

In this study, the cardiac cycle segment data set was obtained through several steps: data cleaning, division, annotation, feature point location and segmentation.

2.1. Data Cleaning

Firstly, the patient medical information corresponding to each ECG is found by retrieving the clinical records. Necessary information includes HIS patient ID, age, date, diagnostic code, main diagnosis name, clinical diagnosis, report description, etc.

Secondly, ECG is screened through patient information. The screening conditions are as follows:

- ECG without corresponding medical record was excluded.
- Because CHD screening is more meaningful for young children, we selected ECG of patients under 5 years old, excluding patients over 5 years old.
- There are generally 12 channels for adult ECG, while children's ECG usually has only 9 channels because they are too small to put so many electrodes. A very few ECG with less than 9 leads were excluded.
- Surgery usually causes ECG changes, which have nothing to do with cardiac structural abnormalities. In order to avoid the influence of this change on CHD screening, ECG during and after operation was excluded.
- For ECGs of the same patient, only the latest ECG was retained.

After multiple rounds of selection, the study cohort eventually consisted of 72626 patients, each with one ECG record.

2.2. Data Division

14% of the cleaned data is used for testing the model, 86% for training the model, and 10% in the training set is used for verification in the training process. The process of data cleaning and division is shown in figure 1.

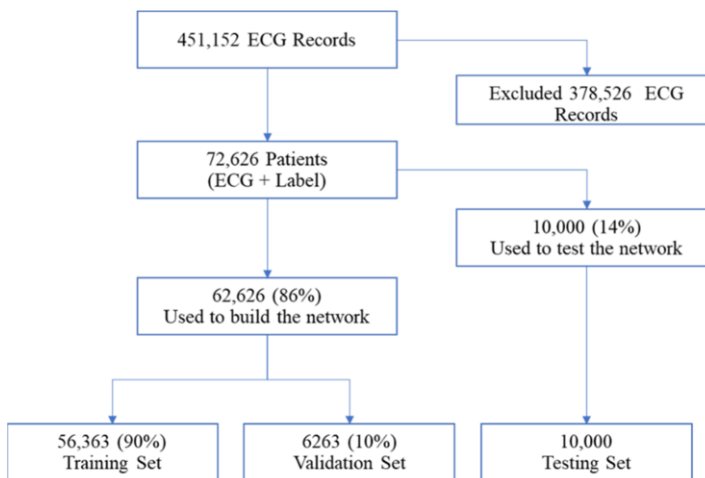


Figure 1. Data cleaning and division.

2.3. Data Distribution and Annotation

The age distribution of 72626 patients is shown in figure 2. The samples aged 2-24 months account for 54% of the total.

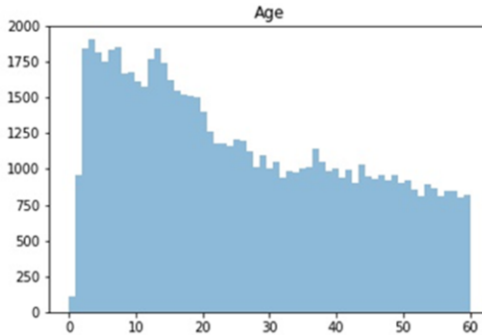


Figure 2. Age distribution of samples.

We use the diagnostic code of the patient's medical record corresponding to the ECG records as the ECG classification label. The corresponding ECG encoded as Q20-Q26 is classified as CHD, and other are classified as non-CHD. After being confirmed by cardiologists, ECG category labels are formed. Of the 72626 samples, 13584 were CHD and 59042 were non-CHD. There are many subtypes of CHD, and the distribution according to its diagnostic code is shown in figure 3. The largest number is Q21, the corresponding main diagnosis name is ventricular septal defect (VSD).

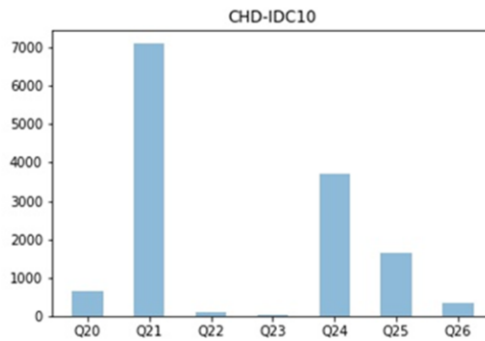


Figure 3. Distribution of CHD subtypes.

2.4. Data Denoising and Feature Location

Firstly, nine ECG channels are intercepted and denoised. Due to the small size of young children, most ECG in the analysis data set only has 9 channels (I, II, III, AVR, AVL, AVF, V1, V3, V5). Therefore, a small number of ECG with 12 channels are also abandoned the other 3 channels (V2, V4, V6). In addition, the signal denoising method based on a hierarchical threshold is used to denoise all ECG data [7].

Secondly, the feature point R are located.

ECG data continuously record the changes of electrical activities in multiple cardiac cycles. Each cardiac cycle is composed of the P wave, QRS wave group, T wave, and

other wavebands. The R vertex of QRS is the most obvious feature point. We developed a positioning algorithm to identify the R point. The characteristic points R of the nine ECG channels are shown by the red dot in figure 4. The number next to the red dot is the index of the point in the array.

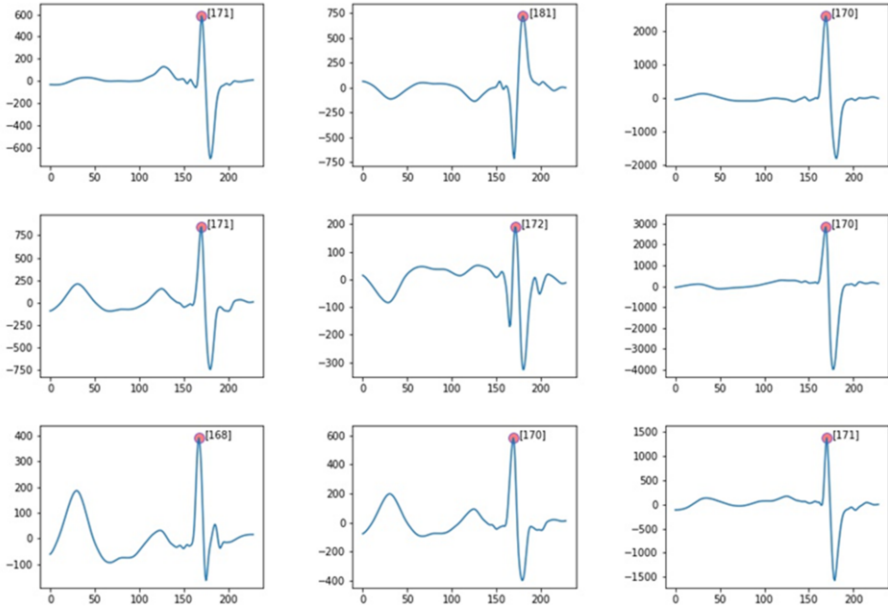


Figure 4. An example of characteristic points of nine ECG channels.

In the example, the abscissa positions of the positioning points of the nine channels are 171,171,168,181,172,170,170,170 and 171 respectively. We take the median value 171 as the overall feature point position of the ECG, record as R.

2.5. Waveform Segmentation

The number of sampling points (i.e., waveform length) of all ECG segments automatically intercepted by the algorithm is between 100-2530. The length distribution is shown in figure 5: blue is the test set, yellow is the training set. We can see that the distribution of the testing set and the training set is basically the same.

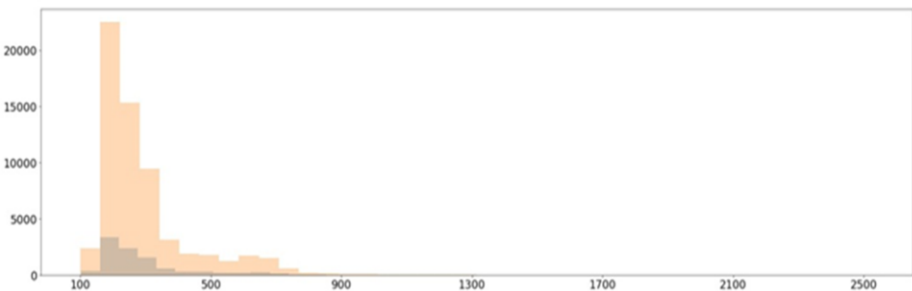


Figure 5. Length distribution of ECG fragments.

Among them, there are 61175 ECG with length between 100 and 343, accounting for 84% of the total. Considering the characteristics of children's heart rate range and depth model training, we selected the segmentation length as 256.

For the waveform length $L < 256$, the two ends are symmetrically complemented, and $(256-L) / 2$ points are supplemented respectively, and the value is the value of the original endpoint.

For the waveform length $L > 256$, take 128 sampling points on the left and right with R value as the center. If there are less than 128 points on the right side of the R, i.e. $(L-R) < 128$, then take $(L-256) \sim L$ point. And the left side is the same.

In this way, each ECG is processed into a two-dimensional array of 9×256 . "9" is the channel number and "256" is the sample number of each channel. Each ECG contains at least one complete cardiac cycle.

3. The Model and Training

We use a Residual of Residual (RoR) network to identify CHD patients which has been described in detail in part III of references [6, 8]. This is an upgraded residual network with two layers of shortcut. The deep network model was implemented with Keras 2.3.1 using TensorFlow backend, and trained on NVIDIA440.44-CUDA10.2-CUDNN7.6.0 in Ubuntu16.04.4. The training set was used to train the deep network (RoR) with the Adam23 optimizer, categorical_crossentropy as the loss function. The internal validation data set was used to tune the hyper-parameters and to select the optimal model.

4. Results and Analysis

Respectively training the model with the whole ECG record data set and the cardiac cycle segment data set, the test results on the independent test set are shown in table 1. They used the same model structure and the same division of patient samples.

It can be seen that the performance of the model after cardiac cycle segmentation has been greatly improved, especially the sensitivity which is our most concerned indicator. The Sensitivity has increased from 0.7473 to 0.8065, which is equivalent to that we can find 6 more potential CHD patients per 100 children, which is of great significance to these children.

Table 1. The performance in the independent testing set.

Model	ECG record	Cardiac cycle segmentation
Test Accuracy	0.9245	0.9443
Precision	0.8462	0.8965
Recall/Sensitivity	0.7473	0.8065
Specificity	0.9407	0.9544
Missed diagnosis rate	0.2527	0.1935
Misdiagnosis rate	0.0328	0.0225
Confusion matrix	[1452,491] [264,7793]	[1567,376] [181,7876]

5. Conclusion

The contribution of this paper is to verify that the interception of cardiac cycle from ECG records can improve the performance of deep learning model in identifying CHD. Previously, the whole ECG record was used to train the model. The model mainly learned the change characteristics of waveform timing. By training the model on the cardiac cycle data set, the depth model can learn the characteristics of a cardiac cycle waveform. Clinically experienced doctors can give conclusions by observing the unique characteristics of ECG waveform. This is consistent with the conclusion of this paper.

References

- [1] Ngeow A, Tan M G, Choo J, et al. Screening for congenital heart disease in a Singapore neonatal unit[J]. Singapore Medical Journal, 2019.
- [2] Wong KK, Fournier A, Fruitman DS, Graves L, Human DG, Narvey M, Russell JL. Canadian cardiovascular Society/Canadian pediatric cardiology association position statement on pulse oximetry screening in newborns to enhance detection of critical congenital heart disease. Canadian Journal of Cardiology 2017;33(2): 199-208.
- [3] Grosse S, Peterson C, Abouk R, Glidewell J, Oster M. Cost and Cost-Effectiveness assessments of newborn screening for critical congenital heart disease using pulse oximetry: A review. International Journal of Neonatal Screening 2017;3(4): 34.
- [4] Yu Q , Cui Y , Zhao P , et al. Conceptual Design of Screening System for Congenital Heart Disease[J]. Journal of Shanghai Jiaotong University (Science), 2019, 24(5):584-590.
- [5] Y. Gong et al., "Fetal Congenital Heart Disease Echocardiogram Screening Based on DGACNN: Adversarial One-Class Classification Combined with Video Transfer Learning," in IEEE Transactions on Medical Imaging, vol. 39, no. 4, pp. 1206-1222, April 2020.
- [6] Yunmei Du, Shuai Huang, Canhui Huang, Allam Maalla, Huiying Liang, "Recognition of Child Congenital Heart Disease using Electrocardiogram based on Residual of Residual Network," 2020 IEEE International Conference on Progress in Informatics and Computing (PIC), Shanghai, China, 2020, pp. 145-148.
- [7] DU Yunmei, HUANG Shuai, LIANG Huiying. Preprocessing of stream denoising for children's EEG signals based on hierarchical thresholds. China Digital Medicine. 2019 14(3): 88-91.
- [8] Yun-Mei D, Maalla A, Hui-Ying L, Shuai H, Dong L, Long L, Hongsheng L. The Abnormal detection of electroencephalogram with Three-Dimensional Deep Convolutional Neural Networks. IEEE Access, vol. 8, pp. 64646-64652, 2020.