Classification of Aortic Stenosis Using ECG by Deep Learning and its Analysis Using Grad-CAM

Erika Hata, Chanjin Seo, Masafumi Nakayama, Kiyotaka Iwasaki,

Takaaki Ohkawauchi and Jun Ohya

***Abstract*— This paper proposes an automatic method for classifying Aortic valvular stenosis (AS) using ECG (Electrocardiogram) images by the deep learning whose training ECG images are annotated by the diagnoses given by the medical doctor who observes the echocardiograms. Besides, it explores the relationship between the trained deep learning network and its determinations, using the Grad-CAM.**

In this study, one-beat ECG images for 12-leads and 4-leads are generated from ECG's and train CNN's (Convolutional neural network). By applying the Grad-CAM to the trained CNN's, feature areas are detected in the early time range of the one-beat ECG image. Also, by limiting the time range of the ECG image to that of the feature area, the CNN for the 4-lead achieves the best classification performance, which is close to expert medical doctors' diagnoses.

***Clinical Relevance*— This paper achieves as high AS classification performance as medical doctors' diagnoses based on echocardiograms by proposing an automatic method for detecting AS only using ECG.**

I. Introduction

For medical examinations of cardiovascular disease, ECG (electrocardiogram), which is simple and non-invasive, is widely used regardless of the scale of medical institutions. However, medical doctors are required much knowledge and long experience in interpreting ECG. Aortic valvular stenosis (AS) is a severe heart valvular disease. It could lead to heart failure, syncope, or sudden death, but it is difficult even for expert medical doctors to detect AS only using ECG. In general, echocardiography is needed for definitive diagnosis.

In recent years, in rural areas in quite many countries such as Japan, the shortage of medical doctors has got serious. Moreover, the medical examination of AS using echocardiography is almost impossible due to the above-mentioned shortage and/or high cost of echocardiography. It can be said that the actualization of an automatic system for diagnosing AS using ECG is desired.

Conventional related work includes deep learning based methods for detecting arrhythmia using ECG. Rahhal *et al.* [1] developed an efficient and robust network that can classify four kinds of arrhythmia by active learning using features obtained by deep learning. Acharya *et al*. [2] achieved an accuracy of 94% for categorizing five types of arrhythmias. Hannun *et al*. [3] accomplished a detection accuracy of expert medical doctors' level using the home-use device called Zio monitor.

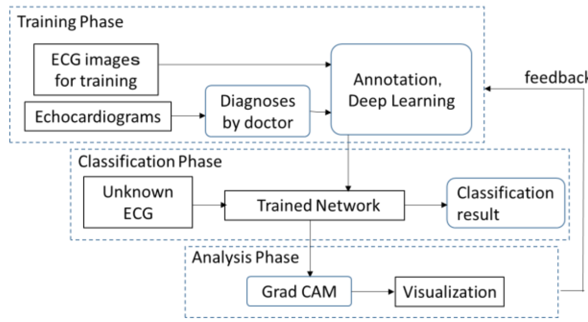

Figure 1. Outline of our proposed methods

Some other studies developed systems that can distinguish the presence or absence of arrhythmia. Goto et al. [4] applied RNN (Recurrent neural network) to 12-lead ECG time-series data to judge the necessity of revascularization surgery for acute coronary syndrome. Also, Attia et al. [5] proposed an inexpensive and noninvasive method for determining asymptomatic left ventricular dysfunction (ALVD). It obtained high false-positive results, but it revealed that the false-positive data correspond to the patient whose probability of developing ALVD in the future is four times as high as the current true-negative person.

However, the aforementioned conventional studies do not clarify how the network generated by deep learning interprets the input and outputs the judgment (diagnosis). Due to differences in human bodies, exceptional data for the network might exist. Even in such a case, it is necessary to be able to verify whether the judgment outputted from the network is appropriate from the medical point of view.

This study focuses on AS, because AS is one of the most serious valvular diseases, and, to the best of our knowledge, no work on the automatic prediction of AS can be seen. This paper proposes an automatic method for classifying AS using ECG by the deep learning network that is trained using ECG training images, each of which is annotated by the medical doctor who performs the echocardiography that corresponds to the ECG image. In this way, echocardiography based on diagnosis is expected to be outputted by inputting only ECG of an unknown patient to the trained network.

Furthermore, to clarify how the trained network interprets the input data and outputs the determination, using the

E. Hata, C.Seo and J. Ohya are with Dept. of Modern Mechanical Engineering, Waseda University, Tokyo, Japan (e-mail: erika.h@moegi.waseda.jp, chanjin@asagi.waseda.jp, ohya@waseda.jp).

M. Nakayama was with Tokyo Medical University, Tokyo, Japan. He is now with the Cardiovascular Centre, Todachuo General Hospital, Toda, Japan (e-mail: masafumi331@gmail.com).

K. Iwasaki is with Cooperative Major in Advanced Biomedical Sciences, Joint Graduate School of Tokyo Women's Medical University and Waseda University, Tokyo, Japan (e-mail: iwasaki@waseda.jp).

T. Ohkawauchi is with the Humanities and Science Department, Nihon University, Tokyo, Japan (e-mail: ohkawauchi@chs.nihon-u.ac.jp).

Gradient-weighted Class Activation Mapping (Grad-CAM) [6]. This paper explores whether the trained network's decision matches the medical doctor's criteria.

II. Proposed Method

As shown in Fig. 1, the proposed method consists of the training phase, classification (prediction) phase, and analysis phase. At first, the training phase trains a Convolutional Neural Network (CNN), and in the classification phase, the trained CNN distinguishes AS or not using ECG of an unknown patient. Then, in the analysis phase, the Grad-CAM is used for analyzing the prediction outputted from the trained CNN, to feedback the analysis to the training phase.

In the training and classification phases, we make 2D waveform images (ECG images) from numeric data for training and classification, respectively. Then, in the training phase, the ECG image annotated by the inspection result given by the cardiologist is inputted to the CNN network, so that the CNN network is trained. In the next phase, the ECG image of an unknown patient is inputted to the trained network so that the network outputs the decision: "AS" or "not AS".

A. Preprocessing

ECG data is stored in the MFER format [7-9], which is a standard for converting general medical waveforms to numeric data, and vice versa. Then, we conduct the following preprocessing based on the literature [10] to the ECG data. The V5 lead, one of the 12-leads of ECG, is used to detect peaks. First, we decompose the signal into coefficients by applying the Wavelet Transform. Second, the coefficients are thresholded to remove the baseline. Third, using the patient's heart rate, peaks of the R wave are detected. Fourth, based on the obtained peaks, the duration from the first to second peaks in the original signal is extracted as a one-beat. Fifth, the one-beat data are temporally re-sampled to obtain 1,000-signal data. Finally, the one-beat ECG image is created as shown in Fig.2, in which the horizontal and vertical directions of the image indicate the signal (time) and potential, respectively. Here, we make the following two kinds of one-beat images, which use all of the 12-leads (Fig.2(a)), and only the four leads: I, aVL, V5 and V6 (Fig.2(b)). The four leads are frequently observed by medical doctors for diagnosing left ventricular hypertrophy, which is common symptoms in AS patients. In the one-beat image, ECG signal values (potential) range between -3 and 3mV.

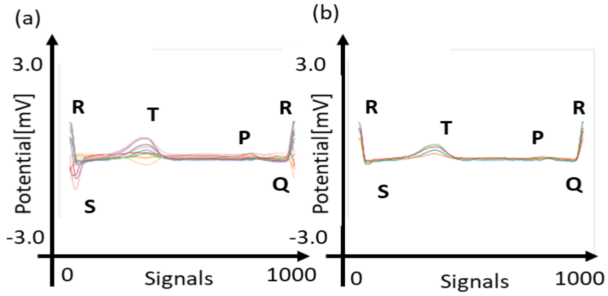

Figure 2. Input images: (a) 12 leads the one-beat image. (b) four leads the one-beat image

B. Network

Each one-beat image for training is annotated by the examination ("AS" or "not AS"), which is based on the guideline of the Japanese Circulation Society [11]. "AS" corresponds to "moderate AS" or "severe AS", and "not AS" corresponds to "functionally normal".

For the CNN network, this paper uses VGG16 [12], because it consists of 16 layers with a 3x3 pixel convolution filter, which leads to high categorization accuracies. We train our network by the fine-tuning whose initial weights are ImageNet's weights, so that reasonable feature extractions are possible despite a small number of training data.

C. Visualize the features by Grad-CAM

We use the Grad-CAM to visualize the gradients of the final convolutional layer of the CNN network in a heatmap. The heatmap can be used for analyzing factors that influence the classification result. Figure 3 (a) shows an example of the obtained Class Activation Mapping (heatmap). The definitions of the horizontal and vertical directions of the map are the same as those of the one-beat ECG image, and the color of each pixel indicates the gradient value (score). The maximum score obtained from the heatmap is used for normalizing the value of each pixel by dividing each pixel's score by the maximal score. Then, the normalized score of each pixel is thresholded by 0.5 so that the feature areas, each of which consists of connected pixels whose scores are 0.5 or greater are extracted. In Fig. 3 (b), the white pixels indicate the pixels included in the feature area obtained from the heatmap in Fig. 3 (a). Finally, each feature area is surrounded by a bounding box (Fig. 3 (c)), and the bounding box is used for the analysis.

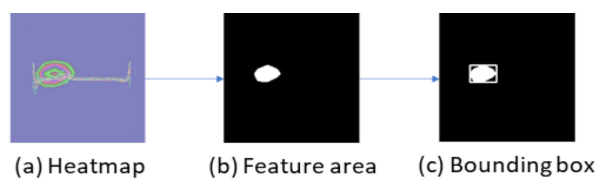

Figure 3. Detecting Feature area using Grad-CAM

III. Experimental Results

A. Dataset and Experiment

Among the 45,478 ECG data recorded from March 29, 2014, to July 10, 2019, at Todachuo General Hospital, we selected 3,513 data whose patients had echocardiograms within three months after taking their ECG. This study involving human subjects was approved by the Institutional Review Board of Todachuo General Hospital. Informed consent was waived because of the retrospective nature of this study. If multiple data of the same person are included, no data or one data is extracted, and the others are excluded. The data of patients using pacemakers are not included in the study. Consequently, the number of ECG data used for the experiments is 700. The data include 108 data of "not AS" and 592 data of "AS". In our study, 108 data of "not AS" are randomly selected. Overall, these data are divided into three groups: 128 for training, 44 for validation, which checks if the weights of the network are good, and 44 for the test. Each group has the same number of "AS" and "not AS" data.

We trained four networks, which are detailed in Sections III.B and III.C. To train these networks with SGD (stochastic gradient descent), we used Keras and three NVIDIA GeForce 1080Ti's. About training conditions, the learning rate was initialized to 0.0001 and batch size was 16.

B. Results of one-beat input networks

We trained two networks using the one-beat image with two different numbers of the lead: 12 and 4. TABLE I shows the average rates of accuracy, recall, precision and F-measure of the test results, where the average rates are computed for the test data. As shown in TABLE I, the evaluations for the 12-lead are better than 4-lead except for the recall rate.

For analyzing how these two networks recognize the input data, we applied the Grad-CAM to both networks. The results are shown in Fig. 4. In each network, a feature area is detected in the left part in the heatmap (early time range) in more than 16 out of the 44 data. The feature area is included in the ST-T part of ECG, where the ST-T part is from the end of the QRS wave to the T wave. As can be seen, the feature areas in the 4-lead (red bounding box) and 12-lead (blue) exist at almost the same position in the ECG image. Each of the two bounding boxes is obtained by averaging the heights and widths of the data in which the feature area could be found. From these results, the position of the two feature areas is expected to play an important role in the classification. Therefore, as described in Section III.C, we build two models whose time range is close to that of the feature areas.

TABLE I. Evaluation of the one-beat input networks

	(i)12-lead ECG	(ii)4-lead ECG
Accuracy	79.5%	77.3%
Recall	72.7%	77.3%
Precision	84.2%	77.3%
F-measure	78.0%	77.3%

C. Result of Input type of Feature Area models

The time (signal) range of the two networks mentioned in Section III.B. is limited between 200*th* and 400*th* signals in the 1000 signals (the normalized duration for one-beat) to focus on the feature area. The signal values of all the input images are normalized so that the signal values range between -1 and 1[mV]. The classification results are listed in TABLE II. According to it, the 12-lead accuracy achieved 81.8%, while the 4-lead accuracy got 88.6%. Overall, these results are better than the results listed in TABLE I. In particular, the 4-lead focusing on the feature area obtained the best results, as shown in TABLEs I and II.

TABLE II. Evaluation of the feature area input networks

	(iii)12-lead ECG	(iv)4-lead ECG
Accuracy	81.8%	88.6%
Recall	81.8%	86.4%
Precision	81.8%	90.5%
F-measure	81.8%	88.4%

From these consequences, it can be said that limiting the time range to the duration for the ST-T feature area and using the 4-lead improves the classification accuracy significantly, compared with the 12-lead with the full-time range (one-beat duration). In case of AS, it is considered that the ST-T part reflects the condition in which the left ventricular is overloaded and blood flow is restricted. Other parts of ECG, such as the R wave, seem not to be essential for the diagnosis.

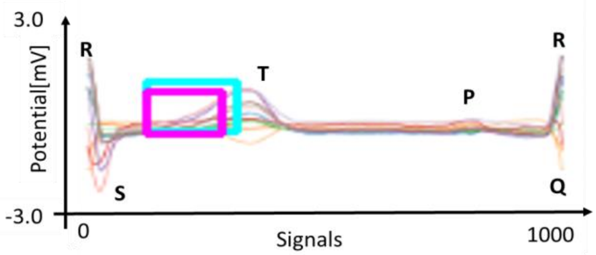

Figure 4. 12-lead and 4-lead one-beat images and their averaged bounding boxes of the feature areas. Blue and red bounding boxes for 12-lead and 4-lead one-beat images , respectively.

IV. DISCUSSION

This chapter explores the relationship between medical doctors' criteria and the judgment results described in Sections III.B and III.C. For this exploration, we need the doctor's criteria for inspecting AS using the echocardiogram. Each country has a diagnostic rule. In Japan, as listed in TABLE III, AS severity is defined according to the following three items; maximum value of aortic flow velocity, mean pressure gradient and aortic valve area by continuity equation [11].

Based on the three items, we analyze FN (false negatives) outputted from two or more of the four models (networks). Examples of such a case are the AS patient data AS_1 to AS_4 in TABLE IV, in which 0 and 1 indicate FN and TP (true positive) predictions outputted from each of the four models, respectively. As shown in TABLE IV, AS_1's judgment is TP in case of 4-lead, but FN in case of 12-lead, while AS_2's is TP in case of one-beat, but not feature area. In contrast, AS_3 and AS_4 data do not give an accurate decision from any model. Here, AS_2, AS_3 and AS_4 in TABLE IV are only three FN data in case of the 4-lead, feature area. It can be seen that these data tend to be misclassified also by the other models (networks).

TABLE III. Diagnostic criteria for degree of AS severity

	mild	moderate	severe
the maximum value of aortic flow velocity [m/s]	< 3.0	$3.0-4.0$	≥ 4.0
mean pressure gradient [mmHg]	< 25	$25-40$	≥ 40
aortic valve area by continuity equation [cm^2]	> 1.5	$1.0-1.5$	≤ 1.0

From TABLE III, the values for the three items in AS_2 to AS_4 correspond to moderate or severe. Therefore, AS_2 to AS_4 can clearly be diagnosed as AS using echocardiography, but are incorrectly classified as shown in TABLE IV. In contrast, AS_1 cannot easily be determined correctly even if echocardiography is used., That is, AS_1 data is close to mild

TABLE IV. AS data for analysis: AS_1 to AS_4 are data whose predictions by two or more models are incorrect. AS_5 and AS_6 are difficult for doctors to judge. 1: TP; 0: FN.

AS patient data ID (identifier)	**AS_1**	**AS_2**	**AS_3**	**AS_4**	**AS_5**	**AS_6**
AS severity	mild - moderate	mild - moderate	moderate	moderate	moderate - severe	severe
the maximum value of aortic flow velocity [m/s]	2.2	3.2	3.7	3.7	2.5	2.6
mean pressure gradient [mmHg]	20	40.7	54	53.4	15.0	27.8
aortic valve area by continuity equation [cm2]	0.9	1.2	0.7	0.8	0.5	0.4
(i) 12-lead, one-beat image	0	1	0	0	1	1
(iii) 12-lead, the feature area image	0	0	0	0	1	1
(ii) 4-lead, one-beat image	1	1	0	0	1	0
(iv) 4-lead, the feature area image	1	0	0	0	1	1

due to the values for the flow velocity and mean pressure gradient, but is severe according to the value for aortic valve area. AS_5 and AS_6 have small values for both flow velocity and mean pressure gradient. Particularly in AS_5 data, two parameters excluding the aortic valve area are mild. In this case, expert medical doctors diagnose, considering other parameters such as skeleton, blood volume and left ventricular ejection fraction. All of the four models predict correctly for AS_5, while for AS_6, correctly except for 4-lead, one-beat.

As described in Section III.B and III.C, our proposed method can accurately classify at high probabilities, which is similar to the doctor's criteria. However, our final goal is to predict accurately, even in such a delicate situation. For this, we still need to collect more various AS data for further analyses.

V. Conclusion

This paper has proposed an automatic method for classifying AS from an ECG image by the CNN trained using the ECG images annotated by the diagnoses given by the medical doctor who observes the echocardiograms. Also, this paper has explored the relationship between trained CNN and its decision using the Grad-CAM.

As a result of training CNN's, by which one heartbeat ECG images are distinguished, and visualizing the trained CNN's by the Grad-CAM, feature areas are detected in the ECG image. By limiting the time range of the ECG image to that of the feature area, better classification results than the one-beat were achieved for 4-lead ECG images. This performance is close to medical doctors' diagnoses based on the echocardiograms.

In the future, we need to collect more AS data so that more accurate classification is possible.

References

[1] M.M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F, Melgani and R.R. Yager, "Deep learning approach for active classification of electrocardiogram signals." Information Sciences, vol.345, July 2016, pp.340-354.

[2] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats." Computers in biology and medicine, vol.89, October 2017, pp.389-396.

[3] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network." Nature medicine, vol.25, January 2019, pp.65-69.

[4] S. Goto, M. Kimura, Y. Katsumata, S. Goto, T. Kamatani, G. Ichihara, S. Ko, J. Sasaki, K. Fukuda and M. Sano, "Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients." PloS one, vol.14, January 2019.

[5] Z. I. Attia, S. Kapa, F. Lopez-Jimenez, P. M. McKie, D. J. Ladewig, G. Satam, P. A. Pellikka, M. Enriquez-Sarano, P. A. Noseworthy, T. M. Munger, S. J. Asirvatham, C. G. Scott, R. E. Carter and P. A. Friedman, "Screening for cardiac contractile dysfunction using artificial intelligence–enabled electrocardiogram." Nature medicine, vol.25, January 2019, pp.70-74.

[6] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra,, "Grad-cam: Visual explanations from deep networks via gradient-based localization." ,IEEE International Conference on Computer Vision., Venice, Italy, October, 2007, pp.618-626.

[7] "Medical Waveform Format Encoding Rules – Part 1", (2006). Retrieved from http://www.mfer.org/doc/MFER_Part1_Ver105.pdf

[8] K. Tamura, M. Hirai, K. Matsumoto, K. Kawamoto, T. Saikawa, M. Horikawa, K. Yamauchi," Medical waveform format encoding rules", ELECTROCARDIOGRAPHY, vol.25, March 2005, pp.151-162.

[9] K. Matsumoto, M. Hirai, H. Kamata, K. Tamura, T. Saikawa, K. Sugi, K. Yamauchi, "Special issue: Past, present and future of MFER", Tokyo Metropolitan of medical technology vol.46, September 2018, pp.160-176 (in Japanese).

[10] Andrew Tan, "Electrocardiograms: QRS Detection Using Wavelet Analysis", medium.com. Retrieved 16 April 2020, from https://medium.com/@andrewtan_36013/electrocardiograms-qrs-detection-using-wavelet-analysis-a1070505efee

[11] The Japanese Circulation Society, (2012). "Guidelines for Surgical and Interventional Treatment of Valvular Haert Disease（JCS 2012）" [PDF file], Retrieved from http://www.j-circ.or.jp/guideline/pdf/JCS2012-_ookita_h.pdf.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Int. Conf. on Learning Representations, San Diego, CA, 2015.