Cardiac detection using YOLO-v5 with data preprocessing

1st Sungjin Lee
Dept. of software convergence
Soonchunhyang university
Asan, Korea
fijianwa@gmail.com

2nd Jia Xibin
Faculty of information technology, Beijng university of technology
Beijing, China
jiaxibin@bjut.edu.cn

3rd Ahyoung Lee
Dept. of computer science
Kennesaw state university
Marietta, USA
ahyoung.lee@kennesaw.edu

4th Hyo-Wook Gil
Dept. of internal Medicine
Soonchunhyang university cheonan hospital
Cheonan, Korea
hwgil@schmc.ac.kr

5th Sungyeup Kim
Dept. of research
All for one
Asan, Korea
sungyeup.kim@gmail.com

6th Min Hong
Dept. of software engineering
Soonchunhyang university
Asan, Korea
mhong@sch.ac.kr

Abstract—Currently, interest in medical-related deep learning is dramatically increasing. Although this interest in deep learning is widely used in other fields, but it is very effective in medical image processing such as CT and MRI, which takes a lot of time for simple medical tests and analysis. In general, in deep learning using such image processing, it is possible to determine which algorithm is the most efficient by collecting data, preprocessing data, and using various models This paper conducted a research on cardiovascular CT images collected from Soonchunhyang University Hospital in Korea, and all of them used data collected for 3 years by professional medical staff. In the case of medical data, the number of data is very limited, so the results can vary greatly depending on how it is processed. Therefore, in this paper, research on an efficient deep learning method was conducted through image data preprocessing using Yolo.

Keywords—deep learning, image processing, YOLO, cardiac, CT, data augmentation

I. INTRODUCTION

Due to economic and technological developments around the world, many people are starting to show a lot of interest in health, which is causing the global population to grow. As the population grows, various health-related medical tests and patients are rapidly increasing, but the number of healthcare-related medical workers is very scarce.. To address these issues, research has been conducted on how to replace simple or costly medical tasks. Currently, tests using deep learning [1] and machine learning [2], a lot of research is being done on interpretation and prediction methods. Research on finding and classifying specific patterns through analysis using time series data, research on text data-based correlation analysis with 7 types of clinical information and results, and research on image-based data learning and prediction such as Computed tomography(CT), Magnetic resonance image(MRI), Positron emission tomography(PET), mammography, ultrasound, and X-ray are typical [3]. In this paper, we are focusing about image-based learning among them. Deep learning algorithms have been developed very rapidly through many researches, and their performance is also increasing every year. In the case of deep learning, an artificial neural network is implemented to mimic the functions of the human cerebral cortex, and the layer of the deep neural network extracts various features and provides several levels of abstraction [4]. However, in the case of DNN, one-dimensional data is basically used. If an image is an input value, spatial/regional information of the image is lost if one line of data is created in the flattening step [5]. Therefore, there is a representative Convolution Neural Network (CNN) specialized for images, and since the image is received as it is, the layer of characteristics is built up while maintaining spatial/local information. In image learning using CNN, we try to perform fast learning using YOLO V5 [6]. Data plays an important role in such image-based learning. However, for medical data, various medical devices and medical software exist, and the results are different. For this reason, there are clear limitations in the process of collecting, so the number of uniform learning data is limited. In order to solve this problem, an efficient method must first be used with a limited number of data, and related studies are being conducted continuously [7-10]. To this end, in this paper, the most obstructive parameters are removed through the OpenCV filter and pixel control. The removed dataset showed better results than the original data after training. In Section II, data augmentation, data processing, and YOLO V5 used in this paper will be explained. In Section III, the methodology will be discussed. Section IV will analyze the results and Section V will discuss the conclusion of our research.

979-8-3503-2028-2/22/$31.00 ©2022 IEEE
DOI 10.1109/CSCI58124.2022.00024

II. BACK GROUND AND RELATED WORKS

A. Data Augmentation and Data Preprocessing

The collection of medical data is bound to be limited by the number of patients and acceptable hospitals. In order to solve this problem, researches on data augmentation and data preprocessing are being actively conducted [11-13]. In the case of data augmentation, it is a method of increasing the number of data itself by creating data slightly different from the original through data size, rotation, inversion, and movement, etc. and using it for learning. And in the case of data preprocessing, we process the data through image processing. Change the image color or set a breakpoint for pixel values to change the contrast value to make boundaries clearer or remove unnecessary parts to facilitate learning. way to make it easier. In this paper, a study was conducted with CT data collected at Soonchunhyang University Cheonan Hospital in South Korea, and it was investigated whether the pre-processing process could make learning easier.

In the case of data to be confirmed in this paper, it is chest CT data to detect cardiovascular disease. Cardiovascular disease is calcified disease, and the proportion of the elderly is increasing as the average life expectancy increases due to the economic and technological development of the population. In proportion to this, a large number of geriatric diseases are also occurring, and cardiovascular diseases occupy a large proportion in these geriatric diseases. According to the World Health Organization (WHO) report, about 17.9 million people died caused by cardiovascular disease (CVD) in 2019, and coronary artery disease (CAD) is the main cause of death among these cardiovascular diseases. In the case of calcification through the aorta, it should be carefully checked because it can be caused by the usual eating habits and lifestyle, as well as the diseases you are suffering from, which can lead to heart attacks and strokes as complications [14, 15].

In the case of Figure 1, it shows the chest CT data of one patient, and in this case, it can be seen that the normal heart is not calcified. However, in the case of the right side, the calcification has progressed to some extent, and when viewed on CT, the right ascending artery has turned white. It is said that calcification has progressed. This is a method to solve this problem because, if the average of 56 CTs per person is 56 CTs per day, even 10 people a day need to see 560 CTs, which are usually detected and analyzed manually by medical experts by self-judgment. I chose to run.

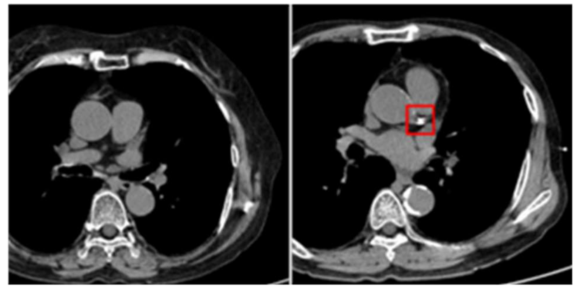

Fig. 1. Normal chest CT(Left), Calcified (Right)

B. YOLO-V5

There are several types of object detection models, and Fig. 2 shows the flow of most detection models. The latest detector mainly consists of two parts: a head and a backbone. The backbone is a part that changes the input image into a feature map, and VGG16 and ResNet-50, which were trained in advance with the ImageNet dataset, are representative. The head is a part that works on the location of the feature map extracted from Backbone, and predict classes and bounding boxes are performed. The head is largely divided into Dense Prediction and Sparse Prediction. This is directly related to whether the object detection is one-stage or two-stage. Two-stage detectors using sparse prediction heads are typically Faster R-CNN and R-FCN. The feature is that Predict Classes and Bounding Box Regression are separated. Representative One-Stage Detectors using Dense Prediction head include YOLO and SSD. Unlike Two-Stage Detector, One-Stage Detector is characterized by integrating Predict Classes and Bounding Box Regression. The neck is the part that connects the head and the backbone, and the feature map is refined and reconstructed. Representative examples include Feature Pyramid Network (FPN), BiFPN, Path Aggregation Network (PAN), NAS-FPN, and the like. The model to be used in this paper is YOLO (You Only Look Once), which is a model widely used in the field of object detection.

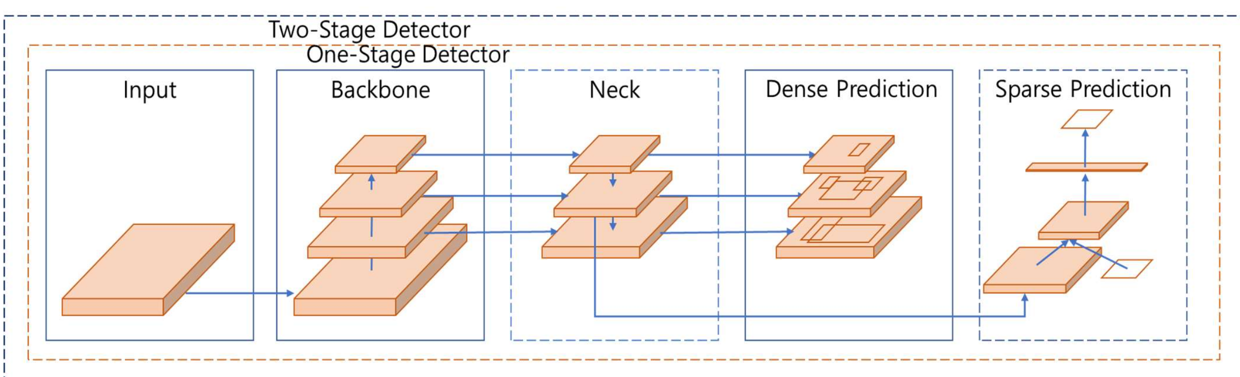

Fig. 2. Architecture of object detection models

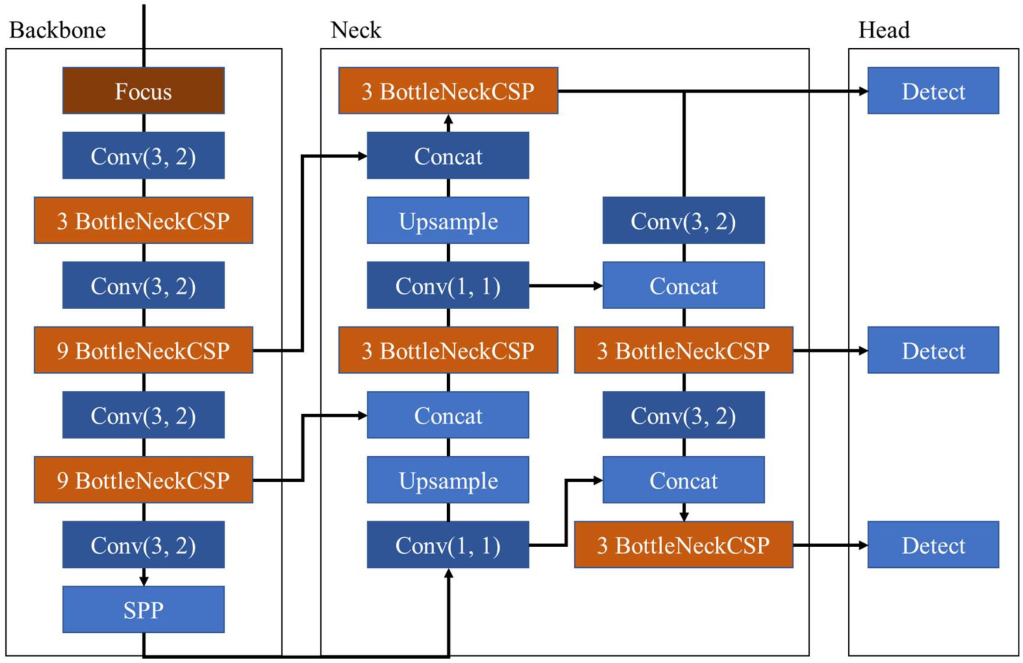

Fig. 3. Composition and flow of YOLO-V5's Backbone, Neck and Head

The characteristic of YOLO is that the entire image is viewed only once, and in the case of R-CNN, the image is divided into multiple images and the image is analyzed using the CNN model.

This has the disadvantage that it takes a lot of time due to the huge amount of computation. On the contrary, YOLO has the characteristic of viewing the image only once without the sphere process. Thanks to these features, YOLO can detect objects in real time. There are several versions of YOLO as well. In the case of the previous version, YOLO V4, CSP-Darkent53 for Backbone, SPP, PAN for Neck, and YOLO V3 for Head. In YOLO V5, SPP() was replaced with SPFP, which is mathematical This is because it produces the same result with a higher speed, but the C3() layer is replaced with a structure that repeats BottleNeckCSP(). In other words, it was replaced to build the model deeper. Fig 3. Shows the architecture of YOLO V5.

III. METHOLOGY

In the case of data augmentation used in this paper, the turning radius was not very large. If the turning radius increases, the complexity of calculations that may occur during learning may become too large, and since the inclination of a person is not so great when performing the original CT scan, an arbitrary inclination was set from -10 degrees to 10 degrees, and the size of the heart depends on the age. , gender, weight, and nutritional status may differ, but it is usually about the size of an adult fist, 325 gm for men and 275 gm for women [16]. So, in the case of data size, I tried changing the size from -10% to 10%, and considering the case where there is noise in the data, I also made data with 2% noise on purpose.

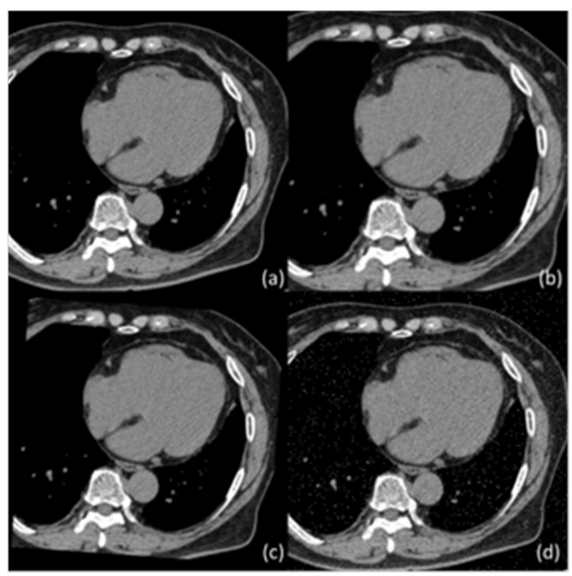

Fig. 4. (a) original data, (b) data with increased size due to cropping, (c) rotated data, (d) data with noise generated

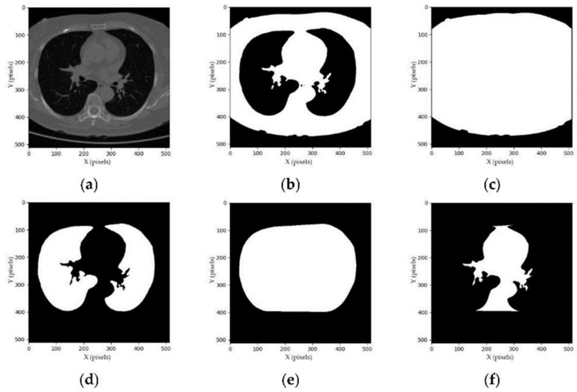

Fig. 5. Flow for erasing the background and other organs

As data pre-processing used in this paper, image processing using OpenCV was performed. First, necessary and unnecessary parts of the data were distinguished from each other. What we need now is only the heart region, so nothing else can make the best data. If we check the data we currently have, it is better to remove it because the ribs and spine have the same pixel values as the aortic lime, and the arteries that are attached to the spine and not the coronary artery are not what we want to see. . Since fat and various organs have parts that are unnecessary for calculation, it is better to remove them, so we thought of an efficient segmentation process for this. First, when CT data was inserted, the data was converted to grayscale and all the pixel values inside the data were changed to white and black to erase the background outside the ribcage. And by designating the internal ROI, it leaves only internal data by filtering substances such as air, fat, muscle and lungs together with the ribs.

From the remaining data, we used Grab cut to separate the spine and heart. The area is divided using the minimal cut algorithm. It is mainly used to distinguish the background from the object in the image by considering the pixel as a graph vertex and finding the optimal cut (Max Flow Minimum Cut) that divides the pixels into two groups. can In the data we currently have, the difference in pixel values between the spine and the heart is clear and used [17]. After preprocessing all data, we trained in YOLO-v5. At this time, batch size was 8, epochs were 300, and SGD was used for the optimizer. During training, 214 layers and 7,022,326 parameters were created.

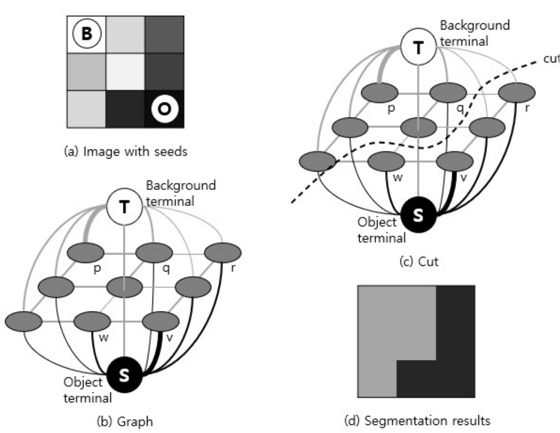

Fig. 6. Flow for Grabcut

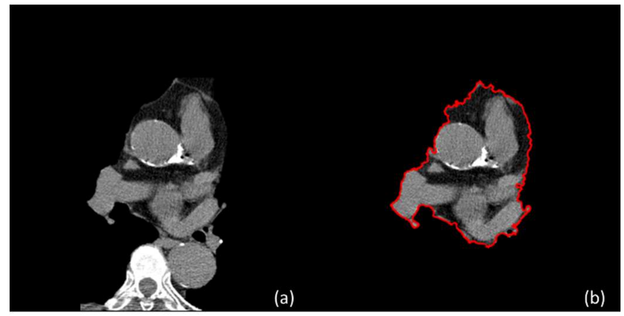

Fig. 7. (a) The result of the algorithm written in Figure 4, (b) Data using Grabcut

IV. Data aquition

The data used in this paper were collected at Soonchunhyang University Cheonan Hospital, passed the Institutional Review Board (IRB), Philips iCT256 was used, and the ISP version was unified to 10.1. The security of the data was thoroughly followed, and all data was anonymized to ensure anonymity. Among the collected data, noise caused by light and a stent treatment that showed a lot of difference from the existing data were treated as exceptions. The original CT data is 512x512 in size, but we changed it to 416x416 in order to reduce the learning time. A total of 600 sheets were used, and the number was increased to 1,800 sheets through data augmentation and used for learning. As shown in Table 1 below, the data set was divided into the ratio of training 7, validation 2, and test 1.

TABLE I. Dataset

	Data type			
	Training	*Validation*	*Testing*	*Total*
CT data	1,260	360	180	1,800

V. Result

In this paper, in order to reduce the complexity and parameters of learning through preprocessing, the training was conducted on the original data with other additional elements and the learned dataset by excluding the ribs, spine, and other organs. As a result of the process, in the dataset created by

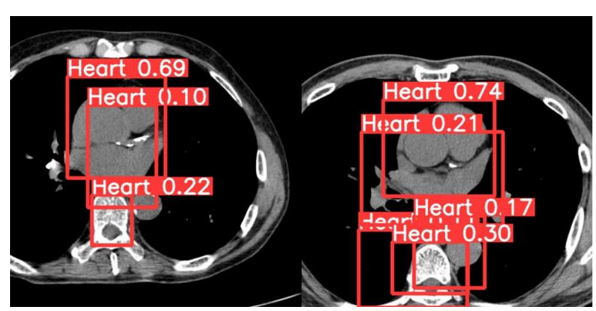

Fig. 8. Errors when training on raw data

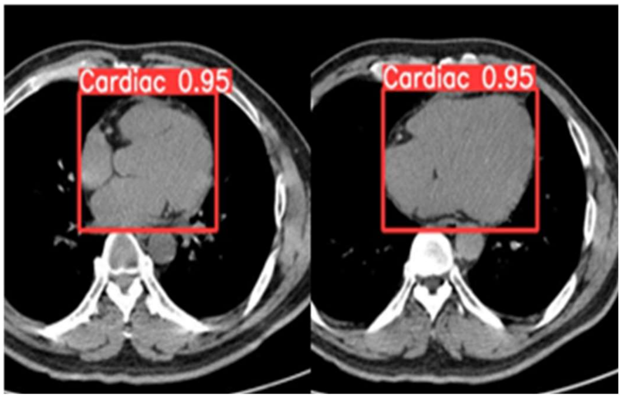

Fig. 9. Result when learning data that has undergone data preprocessing

annotating the original data with the bounding box, as shown in Fig. 8, it was recognized that the heart was recognized in other parts as in the spine or ribs and an error occurred. It was judged as an error that occurred when the learning of surrounding organs and other data was in progress. Through preprocessing, the erased data was learned as shown in Fig. 7, and when the results were viewed, it was confirmed that the results were shown in Fig. 9. Fig. 10 shows the training graph computed with precision and recall. In the case of precision, it means the ratio of correct detection among the detection results, and the recall means the ratio of those predicted to be correct among the actually correctly detected results. As a result, the final result was 0.903 in mAP50-95, which showed a significant difference from the previous original data.

VI. CONCLUSION

In this paper, we conducted a study on how different the learning results can be through the data preprocessing process. Medical data requires a lot of time and money to collect new data, and even if it is attempted through public data and collaboration, data consistency is a little lacking. Therefore, if you directly collect medical data and conduct deep learning, many errors can occur. Accordingly, it is necessary to continue the study of extracting the maximum efficiency with limited data. In this paper, it can be seen that the efficiency is improved by about 40% compared to the case of learning with the original data, which is considered a significant result. Currently, we have conducted on 1,800 pieces of data, but we are aiming for about 5,400 sheets for future research, and although detection is limited to the heart, it is automatically It is expected that the calculation will be completed.

ACKNOWLEDGMENT

This research was supported under the framework of international cooperation program managed by the National Research Foundation of Korea(NRF-2019K1A3A1A20093097) and supported by National Key Research and Development Program of China(2019YFE0107800) and is financially supported by the Ministry of Education(MOE,Korea) through the fostering project of 'Soonchunhyang University and Industry Cooperation Complex' supervised by the Korea Institute for Advancement of Technology(KIAT).

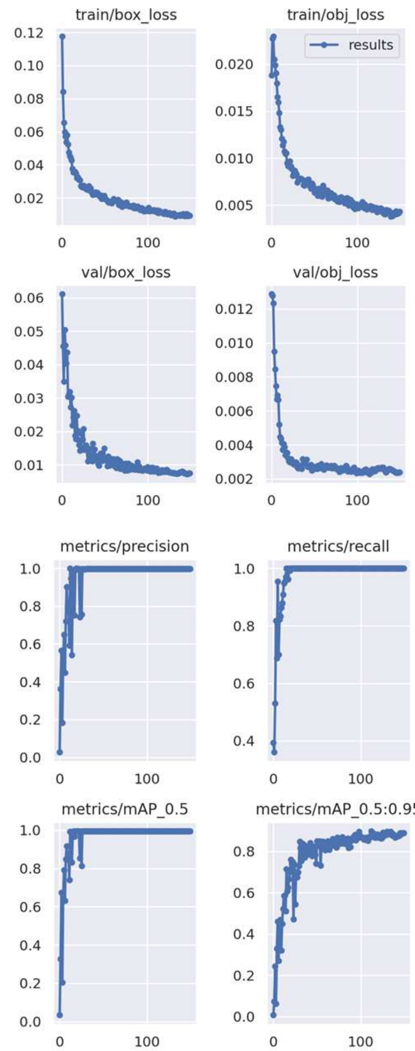

Fig. 10. Result graphs of learning

REFERENCES

[1] Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks. 2015;61:85–117

[2] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260

[3] Brody H. Medical imaging. Nature. 2013;502:S81–S81.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097--1105, 2012.

[5] Zhang, D.; Yao, L.; Chen, K.; Wang, S.; Chang, X.; Liu, Y. Making Sense of Spatio-Temporal Preserving Representations for EEG-Based Human Intention Recognition. IEEE Trans. Cybern. 2019, 50, 3033–3044.

[6] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[7] Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.

[8] Lee, S., Rim, B., Jou, S. S., Gil, H. W., Jia, X., Lee, A., & Hong, M. (2021). Deep-learning-based coronary artery calcium detection from CT image. Sensors, 21(21), 7059.

[9] X. Jia, C. Qian, Z. Yang, H. Xu, X. Han et al., “Boundary-Aware Dual Attention Guided Liver Segment Segmentation Model”. KSII Transactions on Internet and Information Systems (TIIS), vol. 16, no. 1, pp. 16-37, 2022

[10] S. Zhang, C. Y. Pei and W. Y. Liu, "A Real Time Traffic Flow Model Based on Deep Learning." KSII Transactions on Internet & Information Systems, vol. 16, no. 8, pp. 2473-2489, 2022

[11] Desai, S. D., et al. "Deep Neural Network based Medical Image Steganography." 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN). IEEE, 2022.

[12] Xue, Cheng, et al. "Robust Medical Image Classification from Noisy Labeled Data with Global and Local Representation Guided Co-training." IEEE Transactions on Medical Imaging 41.6 (2022): 1371-1382.

[13] Zhang, Ruixuan, et al. "A progressive generative adversarial method for structurally inadequate medical image data augmentation." IEEE Journal of Biomedical and Health Informatics 26.1 (2021): 7-16.

[14] World Health Organization. Cardiovascular Diseases. Available online: https://www.who.int/news-room/fact-sheets/detail/ cardiovascular-diseases-(cvds)

[15] Stary HC. Evolution of atherosclerotic plaques in the coronary arteries of young adults. Arteriosclerosis 1983 ; 3 : 471 a.

[16] Ford, L. E. (1976). Heart size. Circulation Research, 39(3), 297-303.

[17] Allène, Cédric, et al. "Some links between min-cuts, optimal spanning forests and watersheds." ISMM (1). 2007