Ensemble Methods for Estimating the Localization of Coronary Stenosis from CT Images Using 3D CNN Models

Minori Kondo*, Masaki Aono*, Kazuki Shimizu†, Masashi Hahsimoto†, Takeshi Miyaji†, and Kei Nomura†
* Toyohashi University of Technology, Toyohashi, Aichi, JAPAN
E-mail: kondo.minori.kw@tut.jp, masaki.aono.ss@tut.jp
† Toyohashi Heart Center, Toyohashi, Aichi, JAPAN
E-mail: {shimizu, miyaji, k.nomura}@heart-center.or.jp, m.hashimo6123@gmail.com

***Abstract*—In this paper, we focus on accurately detecting and localizing stenosis in non-contrast cardiac CT images. There are different levels of granularity in stenosis localization. We have investigated the localization in several aspects. For instance, consider the aspect of which coronary artery the stenosis is observed and which coronary segments the stenosis has spread to. To shed some light on these localization aspects, we formulate the problem of estimating the location of stenosis in both coronary arteries and coronary artery segments as a multi-label, multi-class classification problem with imbalanced data. We propose ensemble methods with 3D neural network models to predict stenosis localization with high recall. We demonstrate our proposed ensemble approaches through experiments.**

I. Introduction

In recent years, the number of patients with blood vessel-related diseases, such as those affecting the heart and brain, has increased. Cardiovascular hospitals that treat these diseases have large collections of patient images, including computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound. It is challenging to use these images in the medical field to detect diseases early, provide appropriate treatment, and prevent them from becoming severe. The main cause of diseases such as brain diseases and myocardial infarction is generally the accumulation of a substance called plaque in blood vessels. This narrows the blood vessels and causes stenosis. Accurate and early detection of stenosis is truly an urgent issue.

Traditionally, stenosis has been diagnosed visually by radiologists using angiography images during catheter treatment. However, this diagnostic method places a heavy burden on expert radiologists, and the use of contrast agents and highly invasive treatment using catheters places a heavy burden on patients as well. Therefore, we have been conducting research on non-invasive diagnostic techniques that are and less burdensome for patients. One such technique is stenosis prediction from non-contrast CT. In predicting coronary artery stenosis, it is necessary to use AI to assist in estimating the location of stenosis at various levels of detail. This includes not only the presence or absence of stenosis, but also in which coronary artery it occurs. Additionally, it is necessary to determine which segment of which coronary artery is affected.

However, there are several problems that need to be solved in order to estimate the location of stenosis. First, the coronary arteries consist of four groups: RCA, LMT, LAD, and LCX, as shown in Fig. 1. There is also an extreme imbalance in the data regarding where stenosis occurs in the coronary arteries. In fact, even among patients with positive coronary artery stenosis, the number of patients with stenosis in the LMT is extremely low compared to patients with stenosis in other coronary arteries. The same is true for the location of the stenotic segment. There is little data on LMT segment 5, RCA segment 3, or LCX segment 13, while there is an abundance of data for LAD segments 6 and 7. See Fig. 2 for more detail.

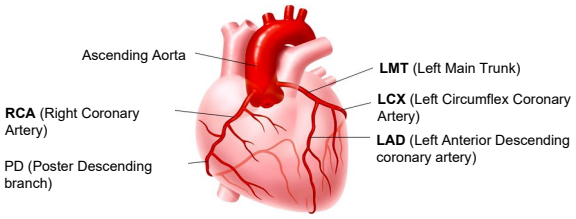

Fig. 1. Illustration of coronary arteries (reproduced and translated into English from https://www.med.jrc.or.jp/hospital/clinic/tabid/146/Default.aspx)

Second, based on our experience, we cannot expect high accuracy in stenosis detection using non-contrast CT images for 2D image classification. Additionally, we have addressed the problem by treating it as a 3D video classification problem in our previous research. Although a 3D neural network model is better than a 2D model, achieving high accuracy remains difficult. Third, in the medical field, it is acceptable to mistakenly diagnose a patient with stenosis and perform a detailed examination when there is no stenosis. In other words, it is acceptable to have a false positive diagnosis of stenosis. However, it is not acceptable to mistakenly diagnose that a patient does not have stenosis when he or she actually does.

In this paper, we address these issues by formulating the problem of estimating stenosis of coronary arteries. We for-

979-8-3315-7206-8/25/$31.00

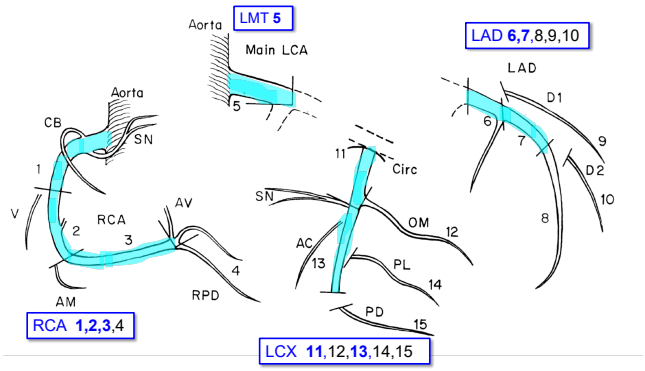

Fig. 2. Illustration of coronary artery segments (annotations added and cited from [1]). Segment numbers in blue represent segments used in our research. Segments numbers in black represent segments not used in our research.

mulate the problem of estimating stenosis of coronary artery segments as a multi-label, multi-class classification problem. We propose ensemble methods using 3D neural network models that can accurately localize stenosis in coronary artery segments. This is achieved despite the imbalance of data between segments. We demonstrate the effectiveness of these methods through experiments.

In Section II, we briefly review related work. Then, in Section III, we describe preprocessing data, followed by a focus on the proposed ensemble methods in Section IV. Section V describes the experimental results. Section VI analyzes our ensemble methods, and Section VII concludes the paper.

II. Related Work

In this section, we briefly describe recent research activities and their objectives related to the prediction of stenosis.

Most approaches to date have employed CCTA (coronary computed tomography angiography), as a reliable, invasive method to assess coronary artery stenosis.

Dan Han et al. [2] used CCTA to assess the coronary artery stenosis, where U-Net [3] was adopted for image segmentation and V-Net [4] was employed for stenosis detection. Andrew Lin et al. [5] used CCTA and ICA (Invasive Coronary Angiography) to predict stenosis.

Abdul Rahman Ihdayhid et al. [6] published their work on assessing coronary artery stenosis and high-risk plaque with a deep learning technique. Specifically they examined approximately 600 patients and translated the problem into four-class classification. Similarly, Jingjing Yan et al. [7] translated the problem into three-class classification.

In our previous work on stenosis severity [8], we defined the problem as a multi-label multi-class classification problem and a regression problem, using 2D DNNs. Similarly, we described our early work on ensemble learning using 3D CNN [9]. Unfortunately in terms of the evaluation criteria, our previous work failed to achieve good performance.

To our knowledge, no other approaches have used cardiac CT slice images to predict coronary artery stenosis, using video-pretrained 3D CNNs, except for our previous research.

III. Overall Flow

Fig. 3 illustrates the overall flow of our proposed system for predicting the stenosis segment. Before we elaborate on each block of the flow, we will take a brief look at the input annotation data and the 3D data representation for the given non-CT images.

A. 3D data representation for CT slices

Among the annotation data to be discussed as shown in Fig. 7, we will focus primarily on the CT slice where the stenosis is most prominent. For *positive* data, we align the center of consecutive CT slices with the lesion number, assuming the location is around the center of the lesion. For *negative* data, we randomly generate a healthy center near the median of all CT slices per subject.

B. 3D block structure

Given a stenosis lesion image within a series of non-contrast CT images, we hypothesize that the adjacent consecutive CT images should also have stenosis. Specifically, for non-contrast CT images, where the average number of CT slices is approximately 56, we hypothesize that six CT slices before and after the central lesion should have stenosis, as shown in Fig. 4. A 3D block consists of $2k+1$ consecutive CT slices: k slices before, one at the central lesion, and k more slices after. After conducting several experiments, we set $k = 6$ in our implementation.

IV. Ensemble Methods

We describe our proposed ensemble methods for predicting stenosis segments from two perspectives. The first, straightforward ensemble method is based on different 3D CNN models. The second ensemble method focuses on one 3D CNN model and combines different loss functions as a weighted sum.

A. Ensemble by different 3D CNN models

The 3D CNN models used in our proposed ensemble methods includes I3D [10], R3D [11], and MC3 [11] models. Fig. 5 illustrates how these models are incorporated into the ensemble. As Fig. 5 shows, fully connected layer ("stacking") learns the weights during ensemble learning.

B. Ensemble by different pretrained loss function parameters

In this approach, we fix one 3D CNN model (here we assume the I3D CNN model), while we combine pretrained models with different loss function parameters. The loss function that we have adopted is focal loss [12]. We chose focal loss for the following reason: It was initially intended to address unbalanced data between classes.

Fig. 6 illustrates the way we did it for an ensemble of different sets of parameters for the focal loss with the I3D 3D CNN model. In our implementation, after several comparative experiments, we set $n = 4$, and adopted the following parameter combinations; $(\gamma_1, \alpha_1) = (0.75, 0.75)$, $(\gamma_2, \alpha_2) = (1.2, 0.6)$, $(\gamma_3, \alpha_3) = (1.5, 0.5)$, and $(\gamma_4, \alpha_4) = (0.5, 0.8)$.

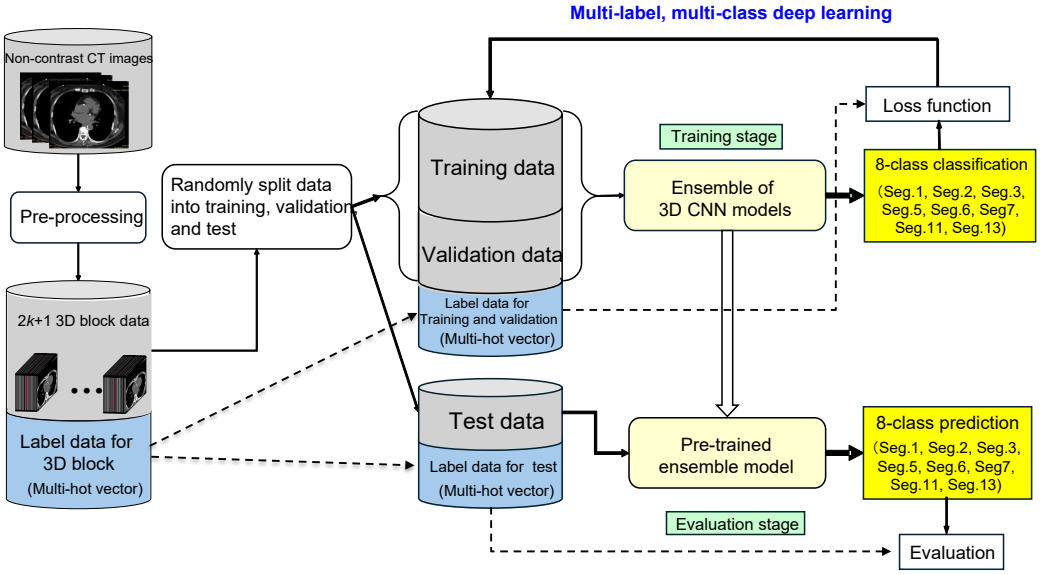

Fig. 3. Overall flow of our proposed system to detect coronary artery segment(s) as a multi-label multi-class classification problem.

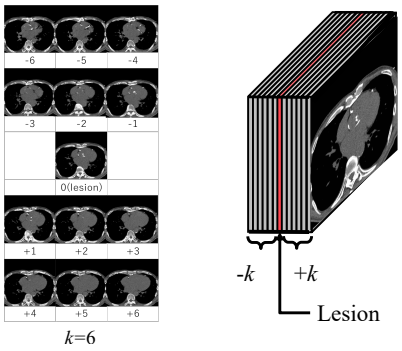

Fig. 4. Representations for non-contrast CT images (a) 2D representation of the $2k+1$ consecutive contrast CT slices, where k slices (in this example, k=6) are hypothesized to inherit stenosis before and after the central lesion image (b) 3D representation of $2k+1$ consecutive slices stick together.

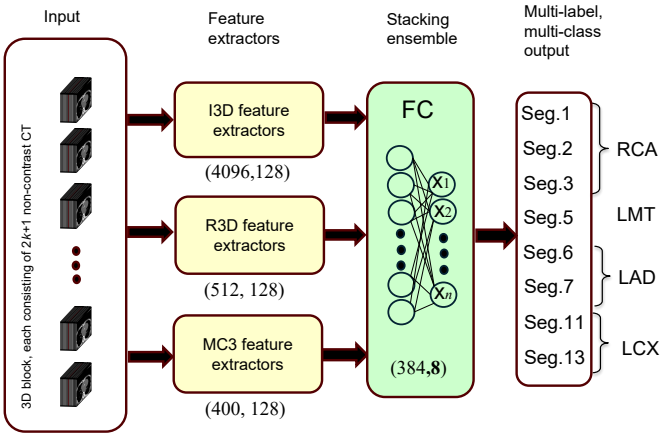

Fig. 5. Ensemble of 3D CNN models. To make each model contribution even, we fine-tune each 3D CNN so that the output dimension of the feature extractor becomes equal (in this example, 128 dimension).

Note that the basic formula of the focal loss is given by the following:

$$FL(p) = -\alpha(1-p)^{\gamma}\log p, \tag{1}$$

where p is probability, α is class balancing factor, and γ is focal parameter.

We carried out extensive experiments and plotted precision, recall, and F1 before determining the above four parameter combinations, including the case of $\gamma = 0.0$, which is reduced to binary cross entropy loss, as well as for $\gamma = 2, \alpha = 0.25$, as recommended by the original authors of focal loss.

In medical diagnosis, recall is highly valued compared to precision. This means that a false positive diagnosis is allowed, indicating that the patients who are diagnosed as stenosis positive can proceed to detailed inspection. Conversely, a false negative diagnosis is not permitted because it means that some patients are diagnosed as negative when they were actually stenosis positive. These considerations lead us to prioritize on recall rather than precision. Therefore, the parameter combination $(\gamma, \alpha) = (2, 0.25)$ should not be permitted, even though this parameter combination leads to the highest precision.

In this section, we will first discuss "annotation data," which includes the diagnostic data of anonymous patients, as shown in Fig. 7. The total amount of stenosis-positive data that we use is approximately 900, while the stenosis-negative data set contains more than 2,000.

C. Annotation data

Fig. 7 illustrates the data that we have received from the Toyohashi Heart Center for stenosis detection, where **No** denotes the patient ID (or number). **CA** indicates the name of coronary artery; RCA, LAD, LCX, or LMT. **segment** is the segment number. **sMLD** is the "simple" lesion number of CT images, referring to the center of the CT slice where stenosis

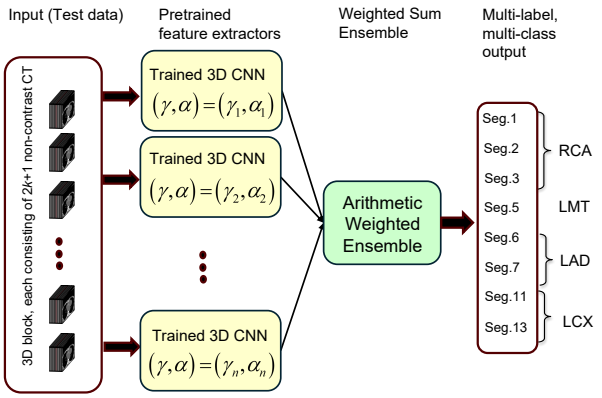

Fig. 6. Ensemble by arithmetic weighted sum of different sets of parameters to Focal Loss with the I3D model

is evident. Note that a **non-contrast** CT image is sometimes referred to as simple, as a subject does not have to ingest a contrast agent prior to the scan.

On average, we obtain 56 non-contrast CT slices, scanned approximately 2.0 mm apart for each subject. **sImage** is a DICOM image that corresponds to the **sMLD**.

No	YN	CA	segment	sLesion	sLesionImage
3	Y	RCA	2	30	0917bb10.dcm
4	Y	LMT	5	7	0918dc06.dcm
6	Y	LAD	6, 7	15	09193f0e.dcm
7	Y	LCX	11	19	09196607.dcm
8	Y	RCA	2	40	0919871c.dcm
8	Y	LAD	6	16	09198704.dcm
10	Y	RCA	1	25	0919f414.dcm
11	Y	RCA	1	13	091a150c.dcm
11	Y	RCA	2	20	091a1513.dcm
11	Y	LAD	7	9	091a1508.dcm
12	Y	RCA	1	22	091a4615.dcm
12	Y	LAD	6	15	091a460e.dcm
13	Y	LAD	7	9	091ab508.dcm

Fig. 7. Annotation data outlook

V. Experimental Results

In this section, we will present the results of our experiments with the proposed ensemble methods and analyze them in terms of recall as an evaluation measure.

Before delving into results, to demonstrate the imbalance of the number of data in coronary arteries, Fig. 8 shows the real data we used in our experiments to demonstrate the imbalance of data in the coronary arteries. Clearly, the coronary artery segments LMT5 and RCA3 are much smaller than the others.

A. 3D CNN Model Ensemble

Fig. 9 shows the recall (in red), precision, and F1 score of the ensemble method using the I3D, R3D, and MC3 models. As shown, the precision is acceptable, while the recall fluctuates among coronary artery segments. Overall, the ensemble achieves 65% recall.

	RCA1	RCA2	RCA3	LMT5	LAD6	LAD7	LCX11	LCX13	healthy	Sum
Training	83	67	27	39	169	170	70	90	392	900
Validation	11	13	5	4	32	33	13	26	60	160
Test	17	18	8	5	27	32	27	32	60	160

Fig. 8. Number of data in each coronary artery

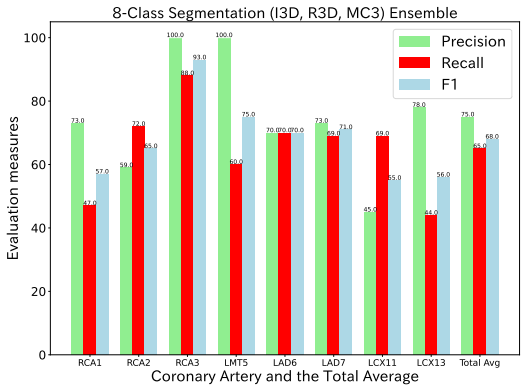

Fig. 9. Precision, recall, and F1 measures of three 3D CNN model ensemble

B. Focal Loss Ensemble

We have conducted experiments by changing (γ, α) parameter combinations as shown in Equation (1).

Fig. 10 demonstrates per coronary artery evaluations.

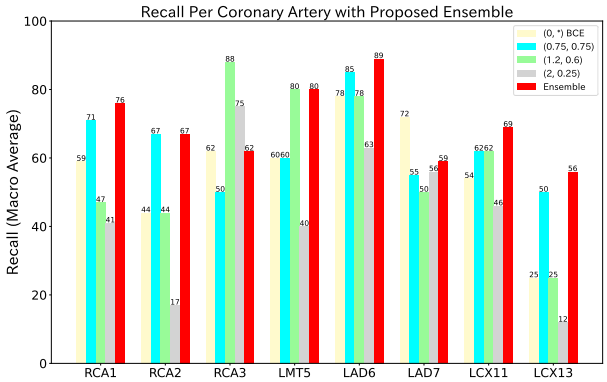

Fig. 10. Recall of each coronary artery with focal loss ensemble method

For comparison, we added several focal loss parameter combinations of (γ, α). Except for RCA3, the recall of the ensemble method was equal to or higher than the others.

VI. Analysis of Ensemble Techniques by Changing the Activation Function Threshold

In our experiments, the output of the ensemble is the eight-class probabilities after applying the sigmoid function. We use a threshold of 0.5, which means if the value exceeds 0.5 then the output becomes 1; otherwise the output becomes 0.

In medical applications, it is usually more important to have a better recall than a better precision. In other words, we want to avoid false negatives, while false positives are acceptable

because they can be subjected to further inspection. Of course a larger recall value is generally favorable, yet this usually comes at the cost of low precision. Thus, we conducted a couple of experiments. We changed the threshold value before the sigmoid activation function to see how it affects the resulting recall and precision.

It should be noted that LMT5 is very important because it is the thickest coronary artery. Thus, if there is positive stenosis in the LMT5, we definitely do not want to miss it.

A. Ensemble of different 3D CNN models

Fig. 11 shows the macro recall and precision curves for an ensemble of three 3D CNN models with varying threshold values. The recall at a threshold of 0.1 is the best. Fig. 12 shows the recall and precision curves for each coronary artery of the ensemble of 3D CNN models. LMT5's recall is invariantly 0.6, regardless of the threshold values.

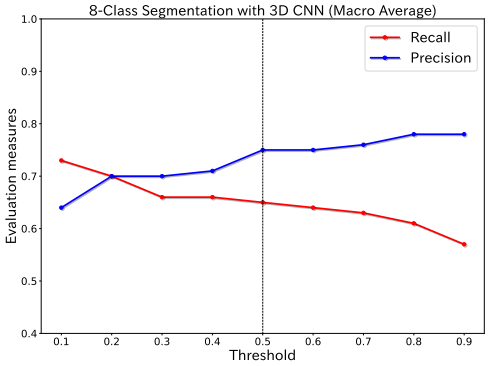

Fig. 11. Precision and recall graphs for the ensemble of three different 3D CNN models

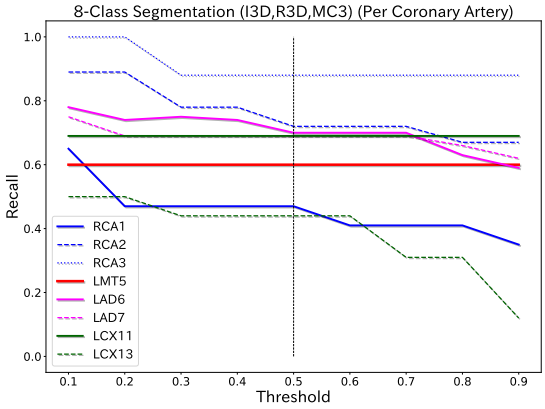

Fig. 12. Precision and recall graphs per coronary artery of the ensemble of three different 3D CNN models

B. Ensemble of CNN models with focal loss parameters

Fig. 13 shows the macro recall and precision curves as threshold values vary. Fig. 14 shows recall and precision curves for each coronary artery. LMT5 has the smallest amount of positive data yet is the most important coronary artery.

According to Fig. 14, the recall of LMT5 is 1.0 when the threshold is less than or equal to 0.2.

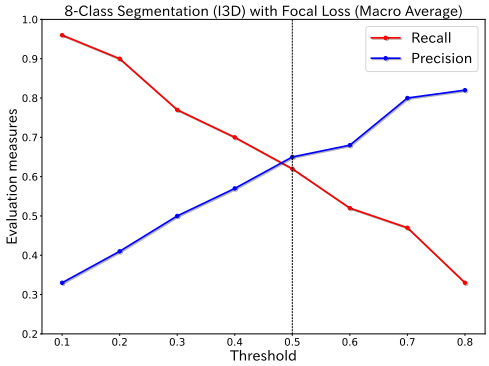

Fig. 13. Precision and recall graphs for the ensemble of four different focal loss parameter settings

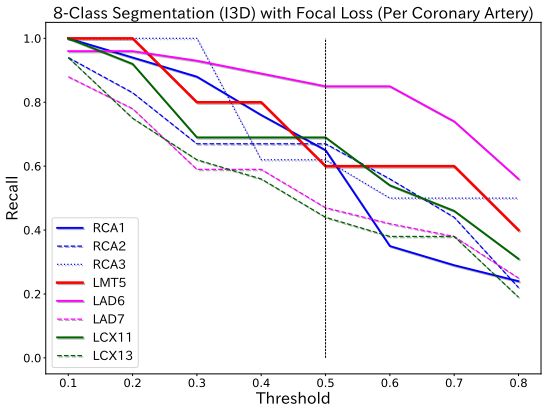

Fig. 14. Precision and recall graphs per coronary artery of the ensemble of four different focal loss parameter settings

C. Summary of Results

Fig. 15 shows the evaluation results of our proposed ensemble methods compared to baseline methods (i.e., methods with a single 3D CNN model). In terms of recall, the ensemble model combined with several focal loss parameters and a threshold 0.2 (equivalent to 1/5) was the best (90%), while in terms of F1 score, the ensemble of 3D CNN models (I3D, R3D, and MC3) with a threshold 0.1 was the best with a score of 68%. Further investigation of this matter is needed.

VII. Conclusions

In this paper, we proposed ensemble methods for estimating the localization of coronary stenosis from non-contrast CT images. Localization in this paper is defined as coronary artery segments. As an important evaluation criterion, we focused on

Model	Precision	Recall	F1	Healthy	Hamming loss	Jaccard Coefficient
I3D	0.68	0.53	0.59	**0.92**	0.10	[0.25,0.33,0.67,0.54,0.53,0.35,0.42,0.31]
R3D	0.54	0.54	0.53	0.83	0.10	[0.35,0.50,0.50,0.08,0.65,0.39,0.40,0.19]
MC3	0.53	0.55	0.50	0.87	0.11	[0.39,0.29,0.28,0.35,0.42,0.37,0.33,0.26]
ensemble (th=1/2) I3D+R3D	0.68	0.60	0.62	0.88	0.09	[0.42,0.50,0.67,0.55,0.52,0.39,0.33,0.32]
ensemble (th=1/2) I3D+R3D+MC3	**0.71**	0.66	0.67	0.87	**0.07**	[0.4,0.44,0.75,0.6,0.51,0.62,0.45,0.30]
ensemble (th=0.1) I3D+R3D+MC3	0.64	0.73	**0.68**	0.88	0.08	[0.44,0.59,1.0,0.43,0.53,0.57,0.33,0.36]
ensemble (th=1/2) focal loss params	0.65	0.62	0.62	0.95	0.08	[0.45,0.54,0.63,0.37,0.48,0.39,0.43,0.32]
ensemble (th=1/5) focal loss params	0.41	**0.90**	0.55	0.62	0.16	[0.42,0.44,0.73,0.19,0.4,0.39,0.33,0.24]

Fig. 15. Summary of Evaluation Results. "th" in Model stands for threshold.

"recall", meaning that we did not want to miss any patients with stenosis.

As the first ensemble method, we described a method that connects multiple 3D CNN models (I3D, R3D, and MC3) with a stacking layer. As the second ensemble method, we described another ensemble method, fixing a 3D CNN model, varying the focal loss parameters. Then, with the pretrained models under four different parameter settings, we introduced an arithmetic weighted sum as an output for predicting stenosis segments. The poposed ensemble methods outperformed methods with a single 3D CNN model.

From our proposed ensemble approaches, we demonstrated that our proposed methods resulted in outstanding recall. This solves the data imbalance problem to some extent among coronary arteries and their associated segments.

Acknowledgment

A part of this research was carried out with the support of the Grant for Toyohashi Heart Center Smart Hospital Joint Research Course.

References

[1] W. Austen et al., "A reporting system on patients evaluated for coronary artery disease. report of the ad hoc committee for grading of coronary artery disease, council on cardiovascular surgery, american heart association," *Circulation*, vol. 51, no. 4, pp. 5–40, 1975. DOI: 10.1161/01.CIR.51.4.5. eprint: https://www.ahajournals.org/doi/pdf/10.1161/01.CIR.51.4.5. [Online]. Available: https://www.ahajournals.org/doi/abs/10.1161/01.CIR.51.4.5.

[2] D. Han, J. Liu, Z. Sun, Y. Cui, Y. He, and Z. Yang, "Deep learning analysis in coronary computed tomographic angiography imaging for the assessment of patients with coronary artery stenosis," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105 651, 2020, ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2020.105651.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241, ISBN: 978-3-319-24574-4.

[4] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, pp. 565–571. DOI: 10.1109/3DV.2016.79.

[5] A. Lin et al., "Deep learning-enabled coronary ct angiography for plaque and stenosis quantification and cardiac risk prediction: An international multicentre study," *The Lancet Digital Health*, vol. 4, no. 4, e256–e265, 2022, ISSN: 2589-7500. DOI: https://doi.org/10.1016/S2589-7500(22)00022-X.

[6] A. R. Ihdayhid et al., "Coronary artery stenosis and high-risk plaque assessed with an unsupervised fully automated deep learning technique," *JACC: Advances*, p. 100 861, 2024, ISSN: 2772-963X. DOI: https://doi.org/10.1016/j.jacadv.2024.100861.

[7] J. Yan et al., "A clinical decision support system for predicting coronary artery stenosis in patients with suspected coronary heart disease," *Computers in Biology and Medicine*, vol. 151, p. 106 300, 2022, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2022.106300.

[8] M. Aono, T. Asakawa, H. Shinoda, K. Shimizu, T. Togawa, and K. Nomura, "Predicting stenosis severity and localization in coronary artery using deep learning," in *Proceedings of the 2024 7th International Conference on Machine Vision and Applications*, ser. ICMVA '24, Singapore, Singapore: Association for Computing Machinery, 2024, pp. 117–125, ISBN: 9798400716553. DOI: 10.1145/3653946.3653964.

[9] M. Aono et al., "Stenosis segment estimation in coronary arteries using ensemble learning of 3d cnn models," in *International Conference on Machine Vision and Applications*, Jun. 2025.

[10] J. Carreira and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017.

[11] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A closer look at spatiotemporal convolutions for action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2018.

[12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017.