Proceedings of International Conference on
Technology and Social Science 2018 (ICTSS 2018)

Investigation of generalization ability by applying 2-Step Learning for Glomerular Epithelial cell images

Yusuke Ono[1, a], Tsutomu Matsuura[1, b], Toshiyuki Matsuzaki[2, c], Keiju Hiromura[2, d] and Takeo Aoki[3, e]

[1]Graduate School of Science and Technology, Gunma University,5-1 Tenjin-cho, Kiryu, Gunma 376-8515, Japan

[2]Graduate School of Medicine, Gunma University, 3-39-22 Showa-machi, Maebashi, Gunma 371-8511, Japan

[3] Gunma Prefectural College of Health Sciences, 323-1 Kamioki-machi, Maebashi, Gunma 371-0052, Japan

[a]<t161b011@gunma-u.ac.jp>, [b]<matsuura@gunma-u.ac.jp>, [c]<matoshi@gunma-u.ac.jp>, [d]<hiromura@gunma-u.ac.jp>, [e]<aoki-take@gch.ac.jp>

Keywords: fisher vector, support vector machine, glomerular epithelial cells

Abstract. We have the impression that Glomerular Epithelial cells are bloating and changing in shape as progress of kidney disease stage. However, these views are only from subjective observations, and so we need to have an objective basis such as statistical method. In order to obtain this basis, we prepared SEM images of Glomerular Epithelial cells which are taken from five types of mice. We divide these images into three groups with a relatively similar tendency. Our target is to catch the changing in shape by classifying these groups with high accuracy. We proposed the classification method with using 2-Step Learning for enhancing the accuracy. In this study, our aim is to clarify problems of this method by investigation of generalization ability. In the result, we confirmed the generalization ability except for certain classes. In contrast, our method has the possibility of enhancing accuracy by improving 1st step.

1. Introduction

Observing Glomerular Epithelial cells, we have the impression that foot processes (Fig.1 Left) are bloating and changing in shape as progress of kidney disease stages. However, these views are only from subjective observations, and so we need to have an objective basis such as statistical method. Thus, we take Glomerular Epithelial cell images from mice in each disease stages. Further, we catch those changes by classifying them with high accuracy. We proposed the classification method with using 2-Step Learning for enhancing the accuracy. In this study, our aim is to clarify problems of 2-Step Learning by investigation of generalization ability.

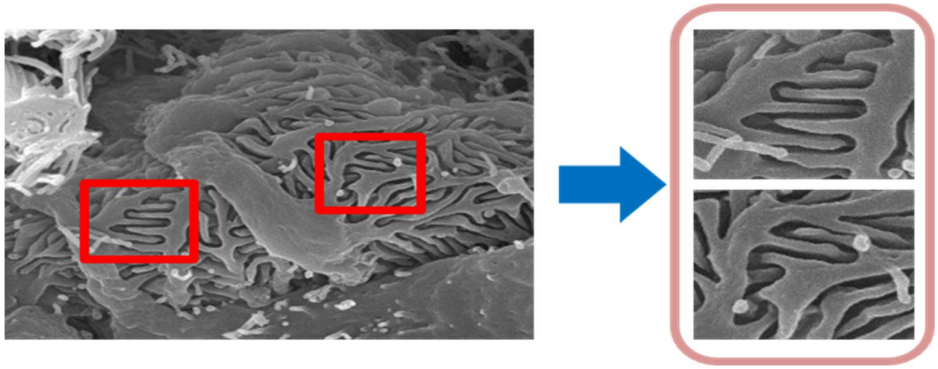

Fig.1 Extraction of Feature Region Images [1]
(Left: Original Image, Right: Feature Region Images(FRI))

2. Image data to be analyzed

We prepared five types of mice that are wild type, wild type with diabetes, aged type, mutant type and mutant type with diabetes. Further, we took foot process regions from them in SEM images. These image data are provided by Prof. Aoki (Gunma Prefectural College of Health Sciences). We divide these images into three groups with a relatively similar tendency, and we classify these three groups. Grouping and the number of each group is shown in Table 1.

Table 1 Grouping and the number of each images

Name	Type	Number(Original)	Number(FRI)
Group1	wild type	96	447
Group2	wild type with diabetes aged type	193	922
Group3	mutant type mutant type with diabetes	167	678

3. About 2-Step Learning

We had thought feature regions that captured an engagement area such as Fig.1 Right expresses features of each group [1]. However, in previous study we achieved the accuracy of classification from about 50 to 60% which was obtained by using those images of the groups. It has low accuracy because features of other classes are sometimes mixed in feature region images obtained from a certain class. Therefore, classifiers should be learned from not only feature region images but also all over images. Hence, we proposed 2-Step Learning method that learns features in two steps. In 1st step, we extract features from feature region images and learn the rough background knowledge. In 2nd step, we obtain more detailed information by scanning and learning the whole of original images with the background knowledge learned in 1st step. In this research, we use FisherVector and SupportVectorMachine as the basis for the feature value and the classifier [2]. In Fig.2-Fig.4, we show the outline of two steps.

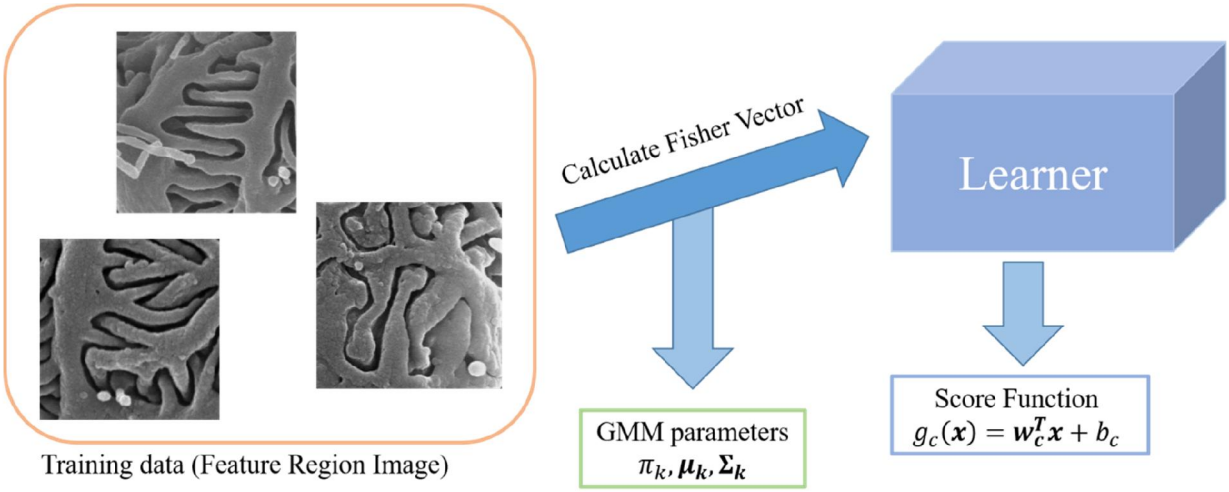

Fig.2 Learning Feature Region Images (1st step)

In 1st step, we learn feature region images and obtain score functions for each group. At this time, we obtain GMM parameters for calculating FisherVector in next step.

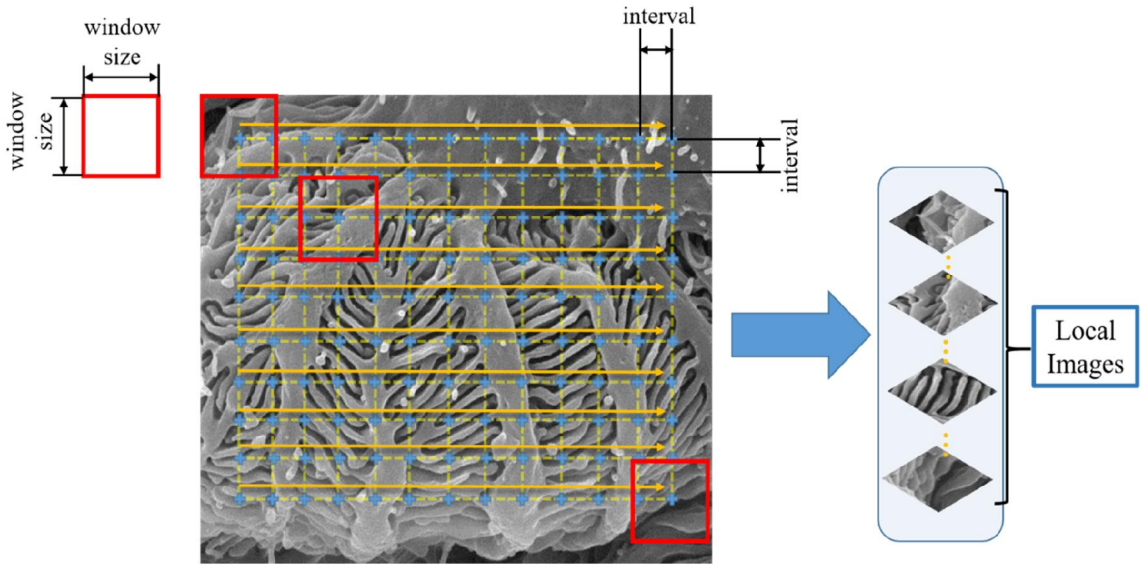

Fig.3 Scanning the whole of original image and obtaining local images (2nd step)

In 2nd step, we scan the whole of original image for obtaining local images. At this time, window size and interval are arbitrarily determined.

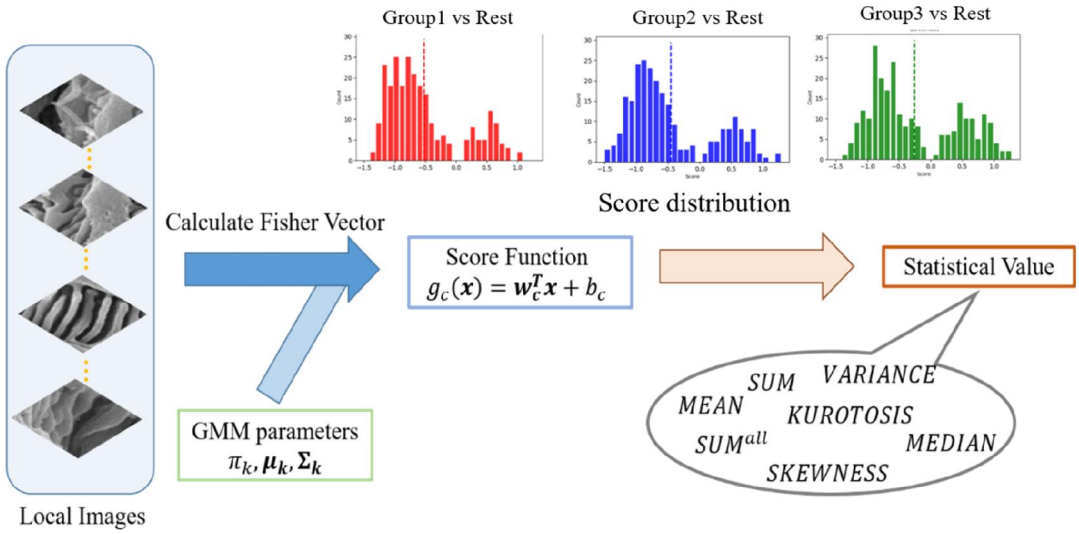

Fig.4 Creating score distribution from local images (2nd step)

We calculate score distribution from local images. Using GMM parameters obtained in 1st step, score value of these images are calculated by Fisher Vector. In addition, we calculate statistical value from the distribution and learn them.

4. Investigation of generalization ability

2-Step Learning achieved about 12.6% higher accuracy of 78.5% than previous research. However, it is unknown whether our method can be applied in general with high precision. Thus, we investigate the generalization ability of our method by exchanging train data and test data at random (Fig.5). In

this time, we exchange these data 10 times and verify it by using mean-accuracy. The result is shown in Table2.

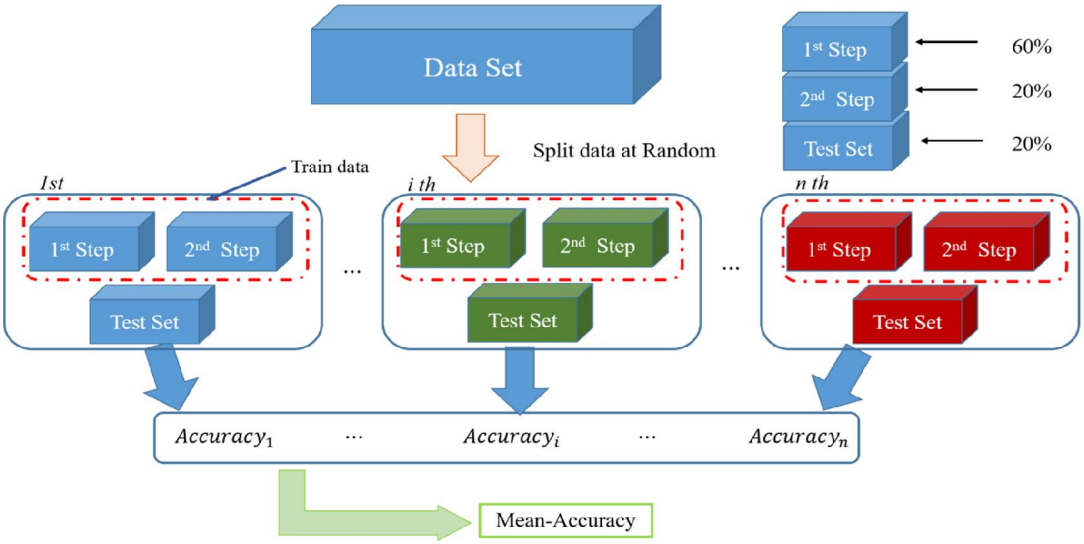

Fig.5 Exchanging train data and test data at random

Table 2 The result of investigation

Name	Mean-Precision [%]			Mean-Accuracy [%]
	Group1	Group2	Group3	
2-Step Learning	21.5	84.7	86.8	72.2

Table2 shows that 2-Step Learning can classify with mean-accuracy of 72.2%. However, mean-precision of Group1 is 21.5%, and this is low value. Depending on the accuracy of learning in 1st step, 2-Step Learning changes the accuracy of score in 2nd step. We think that the leading cause of that phenomenon is from an unsatisfactory accuracy in 1st step.

5. Conclusion

2-Step Learning can classify those groups with mean-accuracy of 72.2%, and we confirm that generalization ability is unsatisfactory in Group1. However, generalization ability of our method has the possibility of enhancing accuracy by improving 1st step.

References

[1] Y. Ono, T. Matsuura, T. Matsuzaki, K. Hiromura and T. Aoki, "Detection of modifications in podocytes and diagnosis of kidney disease stages from SEM of podocyte cells", ***4th International Symposium of Gunma University Medical Innovation, (Maebashi, Japan)***, Nov. 2017.

[2] J. Sanchez, F. Perronnin, T. Mensink, J. Verbeek, "Image Classification with the Fisher Vector: Theory and Practice", ***Research Report n° 8209***, pp4-9, 2013.