Visualization and unsupervised clustering of emphysema progression using t-SNE analysis of longitudinal CT images and SNPs

Hidenobu Suzuki [a], Mikio Matsuhiro [a], Yoshiki Kawata [a], Issei Imoto [b], Yasutaka Nakano [c], Masahiko Kusumoto [d], Masahiro Kaneko [e], Noboru Niki [f]

[a] Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Tokushima, Japan; [b] Division of Molecular Genetics, Aichi Cancer Center Research Institute, Aichi, Japan; [c] Division of Respiratory Medicine, Shiga University of Medical Science, Shiga, Japan; [d] Department of Diagnostic Radiology, National Cancer Center Hospital, Tokyo, Japan; [e] Tokyo Health Service Association, Tokyo, Japan; [f] Medical Science Institute Inc., Tokushima, Japan

ABSTRACT

Chronic obstructive pulmonary disease (COPD) is predicted to become the third leading cause of death worldwide by 2030. A longitudinal study using CT scans of COPD is useful to assess the changes in structural abnormalities. In this study, we performed visualization and unsupervised clustering of emphysema progression using t-distributed stochastic neighbor embedding (t-SNE) analysis of longitudinal CT images, smoking history, and SNPs. The procedure of this analysis is as follows: (1) automatic segmentation of lung lobes using 3D U-Net, (2) quantitative image analysis of emphysema progression in lung lobes, and (3) visualization and unsupervised clustering of emphysema progression using t-SNE. Nine explanatory variables were used for the clustering: genotypes at two SNPs (rs13180 and rs3923564), smoking history (smoking years, number of cigarettes per day, pack-year), and LAV distribution (LAV size and density in upper lobes, LAV size, and density in lower lobes). The objective variable was emphysema progression which was defined as the annual change in low attenuation volume (LAV%/year) using linear regression. The nine-dimensional space was transformed to two-dimensional space by t-SNE, and divided into three clusters by Gaussian mixture model. This method was applied to 37 smokers with 68.2 pack-years and 97 past smokers with 51.1 pack-years. The results demonstrated that this method could be effective for quantitative assessment of emphysema progression by SNPs, smoking history, and imaging features.

Keywords: Radiogenomics, Computed tomography, Single nucleotide polymorphism, t-distributed stochastic neighbor embedding, Emphysema, 3D U-Net

1. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is predicted to become the third leading cause of death worldwide by 2030. A longitudinal study using CT scans of chronic obstructive pulmonary disease (COPD) is useful to assess the changes in structural abnormalities [1]. Previous our study using 5 years follow-up CT images and genomics data showed two single nucleotide polymorphisms (SNPs) associated with emphysema progression [2]. In this study, we performed visualization and unsupervised clustering of emphysema progression using t-distributed stochastic neighbor embedding (t-SNE) [3] analysis. The procedure of this analysis is as follows: (1) automatic segmentation of lung lobes using 3D U-Net, (2) quantitative image analysis of emphysema progression in lung lobes, and (3) visualization and unsupervised clustering of emphysema progression using t-SNE.

2. MATERIALS AND METHODS

2.1 Longitudinal CT image and SNP database

Longitudinal CT images were collected from the Tokyo health service association. Collection and analysis of data were approved by the Institutional Review Board at the institution. The CT images were acquired on Aquilion scanner with 30 mA at 120 kVp, plane resolution: 0.625 or 0.781 mm, reconstruction matrix: 512 x 512, convolution kernel: FC01, slice

Medical Imaging 2022: Computer-Aided Diagnosis, edited by Karen Drukker, Khan M. Iftekharuddin, Hongbing Lu, Maciej A. Mazurowski, Chisako Muramatsu, Ravi K. Samala, Proc. of SPIE Vol. 12033, 120331H · © 2022 SPIE · 1605-7422 · doi: 10.1117/12.2609296

thickness: 1.0 mm, and reconstruction interval: 1.0 mm. The number of smokers and past smokers were 37 and 97, respectively, follow-up years were 5.3+1.5 and 5.3±1.4 years, and pack-years were 68.2±30.0 and 51.1±29.7 pack-years.

2.2 Automatic segmentation of lung lobes using 3D U-Net

The lungs are divided into five lobes; RUL (right upper lobe), RML (right middle lobe), RLL (right lower lobe), LUL (left upper lobe), and LLL (left lower lobe). Excellent works have been presented for lung lobe segmentation [4][5]. Here, we used 3D U-Net [6] for lung lobe segmentation. The network architecture is shown in Fig. 1. We focused on the vascular tree belonging to each lung lobe, and blank spaces between these trees. The input volume was scaled into a small volume to reduce the detail of the tree. The input to the network was 128 x 128 x 128 voxel with a channel. The output of the network was 128 x 128 x 128 voxel with five channels corresponding to RUL, RML, RLL, LUL, and LLL. The 3D U-Net architecture was implemented in TensorFlow [7]. The network was trained with 200 epochs using 270 scans for training and 30 scans for validation, using the Adam optimization algorithm, with the Dice loss function. This training was conducted on a single graphical processing unit (NVIDIA GeForce RTX 3090). The segmentation results of the lung lobe were restored to the original size using linear interpolation. The segmented regions were used as seeds to fill the lungs by the five lobes. The segmentation of lungs was performed using a previous method [8].

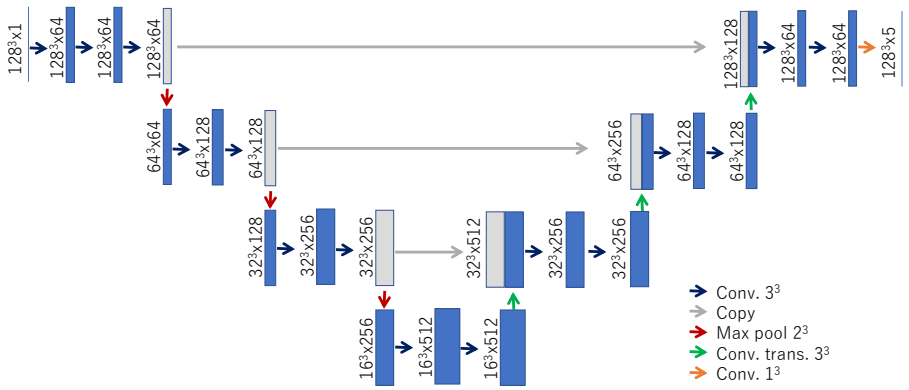


Fig. 1. Network architecture for lung lobe segmentation.

2.3 Quantitative image analysis of emphysema progression in lung lobes

CT can detect the loss of lung tissue associated with emphysema as low attenuation volume (LAV) [1]. LAV% was defined as the percentage of voxels less than a threshold of -950 H.U. [9][10]. Its progression was defined as the annual change in LAV%, which was computed by linear regression of time-series LAV% [11]. Examples of the emphysema progression were shown in Fig. 2. We measured two imaging features to quantify the LAV distribution in lung lobes; LAV size and LAV density. Since the LAV clusters are frequently connecting each other, cluster volume is not

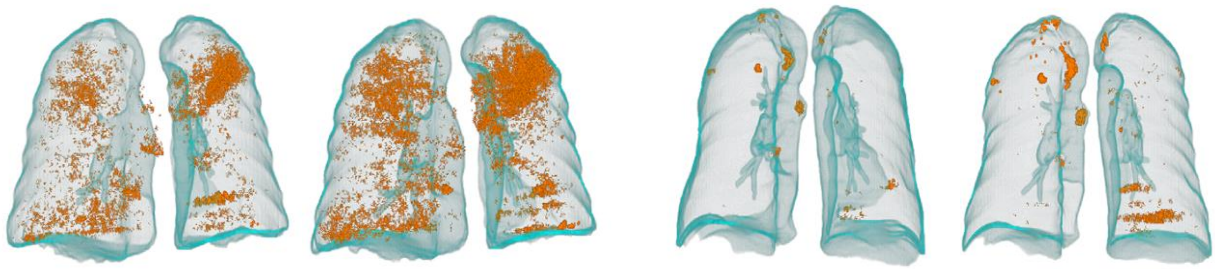


Fig. 2. Quantitative evaluation of emphysema progression for two participants. The annual increments of LAV% for the participant (a) and (b) were 0.31 %/year and 0.05 %/year, respectively.

appropriate to the LAV size. 3D Euclidean distance was employed as LAV size. LAV density was defined by the LAV per $1cm^3$.

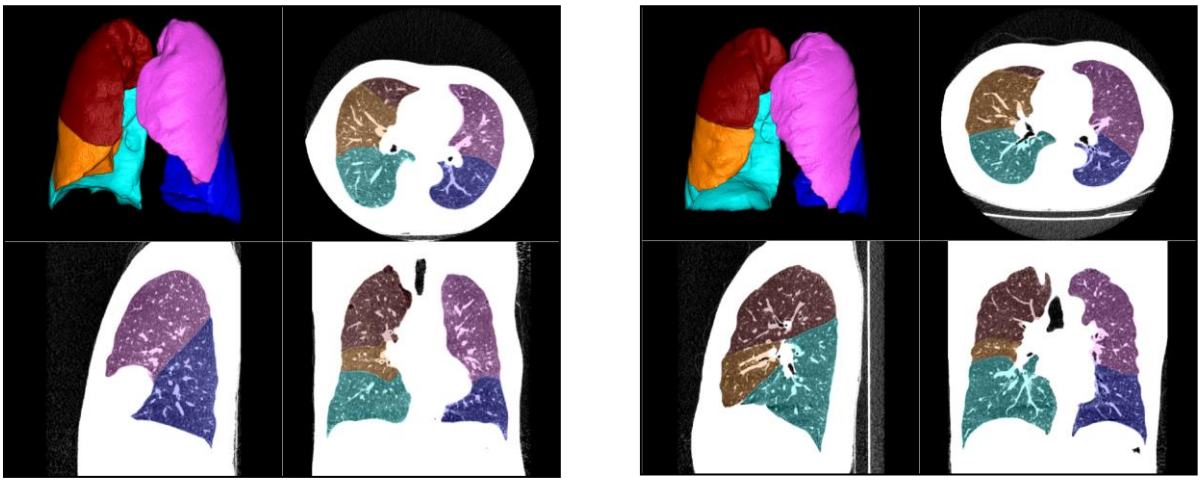
2.4 Visualization and unsupervised clustering of emphysema progression using t-SNE

In our previous study [2], it was presented that two SNPs are associated with emphysema progression; rs3923564 in SFTPD [12] and rs13180 in IREB2 [13]. These genotypes at the SNPs, smoking history (smoking years, number of cigarettes per day, pack-year), and LAV distribution (LAV size and density in upper lobes, LAV size, and density in lower lobes) were combined into nine-dimensional space. This space was transformed to two-dimensional space by t-SNE, and divided into three clusters by Gaussian mixture model. The perplexity parameter and iteration were set to 15 and 2000, respectively.

3. RESULTS

3.1 Lung lobe segmentation results

An observer semi-automatically determined the ground truth of the lung lobe using our previous method which separates lung lobes based on the interlobar fissures detected by four-dimensional curvature features. This method was tested with 32 scans. The performance of segmentation was evaluated according to the Dice similarity coefficient (DSC). DSCs of RUL, RML, RLL, LUL, and LLL segmentation were 0.971 ± 0.011, 0.935 ± 0.089, 0.944 ± 0.172, 0.977 ± 0.003, and 0.973 ± 0.004, respectively. Examples of lung lobe segmentation are shown in Fig. 1.



Participant 1 Participant 2

Fig. 3. Lung lobe segmentation results of two participants. Red, orange, light blue colors show RUL, RML, and RLL, respectively. Magenta and blue colors show LUL and LLL, respectively. Window center and width for CT images are -800 H.U. and 400 H.U.

3.2 Visualization and unsupervised clustering results of emphysema progression using t-SNE

Visualization and clustering results using t-SNE are shown in Fig. 4. A comparison of emphysema progression in three classes is shown in Fig. 5. The mean emphysema progressions for classes 1, 2, and 3 were 0.20%/year, 0.06%/year, and 0.04%/year, respectively. The emphysema progression of class 1 was significantly larger than the other classes (pairwise t-test with Bonferroni correction). Smoking history (smoking years, number of cigarettes per day, and pack-year) of three classes are shown in Fig. 6. There was no significant difference between the pack-year of classes 1 and 3. Genotypes at rs13180 and rs3923564 in the three classes are shown in Table 1. There was a significant difference in the genotypes at rs13180 (Fisher's exact test, p-value<0.05), and not in the genotypes at rs3923564. The size and density of LAV for upper lobes are shown in Fig. 7. Both of size and density for class 1 were significantly larger than the other two classes (p-values<0.05, pairwise t-test with Bonferroni correction). Similar trends were observed for lower lobes. Focusing on class 1, the LAV densities of almost all patients were high. However, LAV sizes look like to be classified into two

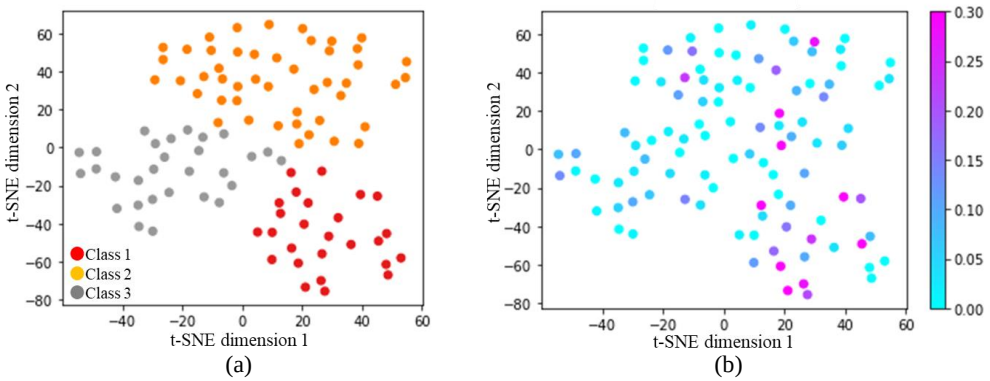


Fig. 4. Visualization and unsupervised clustering results of emphysema progression. (a) Clustering result, (b) emphysema progression (%/year). The mean emphysema progression for class 1, 2, and 3 were 0.20, 0.06, and 0.04, respectively.

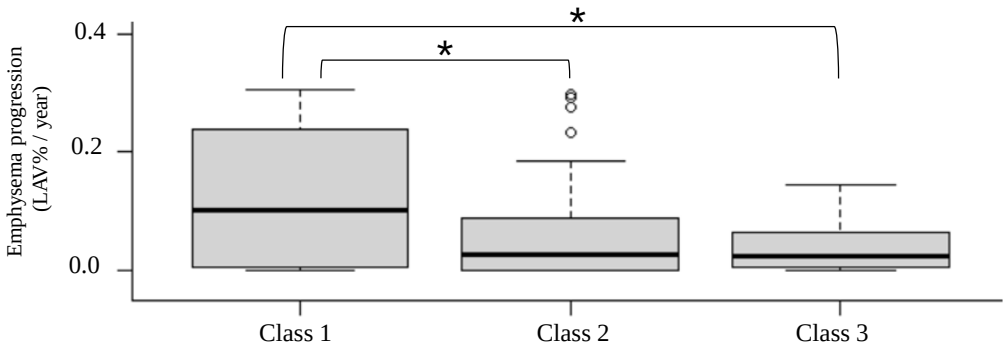


Fig. 5. Comparison of emphysema progression (LAV%/year) in three classes (* *p*-values<0.05, pairwise t-test with Bonferroni correction).

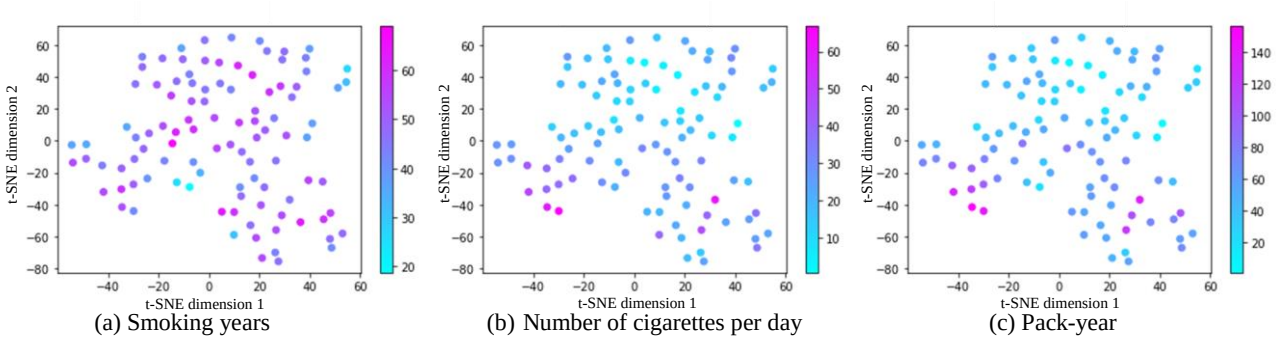


Fig. 6. Visualization of smoking history of the three classes. (a) Smoking years of class 1, 2, and 3 were 49.6, 47.4, and 46.5, respectively. (b) Number of cigarettes per day of class 1, 2, and 3 were 28.3, 15.4, and 30.4, respectively. (c) Pack-year of class 1, 2, and 3 were 69.0, 35.4, 71.1, respectively.

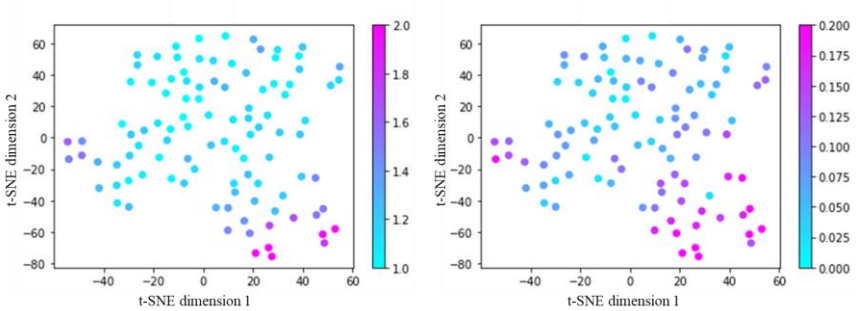
groups. A group has a large LAV size and high density, and another group has a small LAV size and high density. There is a difference in the distribution of the emphysematous regions as shown in Fig. 1. Additional local and global imaging features would be useful information for separating these three classes into subdivided classes.

4. CONCLUSIONS

This paper presented a visualization and unsupervised clustering of emphysema progression using t-SNE analysis of longitudinal CT images and SNPs. The visualization and clustering using t-SNE could be an effective method for quantification of emphysema progression by SNPs, smoking history, and imaging features.

Table 1. Genotypes at rs13180 and rs3923564 of three classes.

	rs13180			rs3923564		
	C/C	C/T	T/T	A/A	A/G	G/G
Class 1	5	18	3	13	9	4
Class 2	4	25	18	18	18	11
Class 3	18	11	0	10	12	7



(a) LAV size (b) LAV density

Fig. 7. Visualization of LAV size and density of the three classes. (a) LAV size of class 1, 2, and 3 were 49.6, 47.4, and 46.5, respectively. (b) Number of cigarettes per day of class 1, 2, and 3 were 28.3, 15.4, and 30.4, respectively. (c) Pack-year of class 1, 2, and 3 were 69.0, 35.4, 71.1, respectively.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 19K12784, and by the SECOM Science and Technology Foundation.

REFERENCES

[1] Takayanagi, S., Kawata, N., Tada, Y., Ikari, J., Matsuura, Y., Matsuoka, S., Matsushita, S., Yanagawa, N., Kasahara, Y., and Tatsumi, K., "Longitudinal changes in structural abnormalities using MDCT in COPD: do the CT measurements of airway wall thickness and small pulmonary vessels change in parallel with emphysematous progression?", Int J Chron Obstruct Pulmon Dis., 12, 551-560 (2017).

[2] Suzuki, H., Matsuhiro, M., Kawata, Y., Niki, N., Imoto, I., Nakano, Y., Kusumoto, M., and Kaneko, M., "Association analysis of SNPs with CT image-based phenotype of emphysema progression in heavy smokers", Proc SPIE 11314, 113142D (2020).

[3] Van der Maaten, L., and Hinton, G., "Visualizing data using t-SNE", Journal of machine learning research, 9(11) (2008).

[4] Xie, W., Jacobs, C., Charbonnier, J. P., van Ginneken, B., "Relational modeling for robust and efficient pulmonary lobe segmentation in CT scans", IEEE Trans Med Imaging, 39(8), 2664-2675 (2020).

[5] Park, J., Yun, J., Kim, N., Park, B., Cho, Y., Park, H. J., Song, M., Lee, M., and Seo, J. B., "Fully automated lung lobe segmentation in volumetric chest CT with 3D U-Net: Validation with intra- and extra-datasets", J Digit Imaging, 33, 221-230 (2020).

[6] Çiçek, Ö., Abdulkadir, A., Lienkamp, SS., Brox, T., and Ronneberger, O., "3D u-Net: Learning dense volumetric segmentation from sparse annotation", International Conference on Medical Image Computing and Computer-Assisted Intervention, 424-432 (2016).

[7] Abadi, M., Barham, P., Chen, J., Chen, Z, Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P,., Wicke, M., Yu, Y., and Zheng, X., "Tensorflow: A system for large-scale machine learning", 12th USENIX symposium on operating systems design and implementation, 265-283 (2016).
[8] Matsuhiro, M. Suzuki, H. Kawata, Y. Niki, N. Nakano, Y. Ohmatsu, H. Kusumoto, M. Tsuchida, T. Eguchi, K. Kaneko, M. "Peripleural lung disease detection based on multi-slice CT images," Proc. of SPIE, 9414, 94142W (2015)
[9] Gevenosis, P. A., de Maertelaer, V., de Vuyst, P., Zanen, J., and Yernault, J., "Comparison of Computed Density and Macroscopic Morphometry in Pulmonary Emphysema," Am. J. Respir. Crit. Care. Med. 152, 653-657 (1995).
[10] Parr, D. G., Stoel, B. C., Stolk, J., Nightingale, P. G., and Stockley, R. A., "Influence of Calibration on Densitometric Studies of Emphysema Progression Using Computed Tomography," Am. J. Respir. Crit. Care Med. 170(8), 883-890 (2004).
[11] Suzuki, H., Mizuguchi, R., Matsuhiro, M., Kawata, Y., Niki, N., Nakano, Y., Ohmatsu, H., Kusumoto, M., Tsuchida, T., Eguchi, K., Kaneko, M., and Moriyama, N., "Quantitative assessment of smoking-induced emphysema progression in longitudinal CT screening for lung cancer", Proc. SPIE 9414, 94142O (2015).
[12] Zhou, H., Yang, J., Li, D., Xiao, J., Wang, B., Wang, L., Ma, C., Xu, S., Ou, X., and Feng, Y., "Association of IREB2 and CHRNA3/5 polymorphisms with COPD and COPD-related phenotypes in a Chinese Han population", Journal of Human Genetics, 57, 738-746 (2012).
[13] Chen, W., Brehm, J. M., Manichaikul, A., Cho, M. H., Boutaoui, N., Yan, Q., Burkart, K. M., Enright, P. L., Rotter, J. I. Petersen, H., Leng, S., Obeidat, M., Bossé,, Y., Brandsma, C., Hao, K., Rich, S. S., Powell, R., Avila, L., Soto-Quiros, M., Silverman, E. K., Tesfaigzi, Y., Barr, R. G., and Celedón, J. C., "A Genome-Wide Association Study of Chronic Obstructive Pulmonary Disease in Hispanics", Annals of American Thoracic Society, 12(3), 340-348 (2015).
[14] Matsuhiro, M., Suzuki, H., Kawata, Y., Niki, N., Ueno, J., Nakano, Y., Ogawa, E., Muro, S., Mishima, M., Ohmatsu, H., and Moriyama, N., "Extraction method of interlobar fissure based on multi-slice CT images", Proc SPIE 8670, 867031 (2013).