Comparison of Feature Selection method for Diffuse Lung Disease

Satoshi ONO[1], Makoto KOIWAI[2], Hayaru SHOUNO[2], Shoji KIDO[3]
shouno@uec.ac.jp
[1] Department of Informatics and Engineering, University of Electro-Communications, Chofugaoka 1-5-1, Chofu, 182-8585, Tokyo, JAPAN
[2] Graduate School of Informatics and Engineering, University of Electro-Communications, Chofugaoka 1-5-1, Chofu, 182-8585, Tokyo, JAPAN
[3] Graduate School of Medicine, Yamaguchi University, Tokiwadai 2-16-1, Ube, 755-8611, Yamaguchi, JAPAN

Abstract— *Diffuse lung disease (DLD) in high resolution computed tomography images show a lot of variations even in the same class, and this variations make difficulty in diagnosis. For effective diagnosis, such like classification of disease, feature selection is a important factor. In the field of pattern classification, a lot of feature selection methods have been proposed. The feature selection is a kind of combinatorial problem, so that it is hard to solve. In this study, we compare several methods from the classification viewpoint. We prepare the conventional method, the sparse modeling methods, and the method based on the Markov-Chain Monte Carlo (MCMC) method. In the result, the modern feature selection method using sparse modeling shows better result rather than that of the conventional methods. The sparse modeling based methods looks close up to the one with MCMC method, which is an approximation version of brute-force searching.*

Keywords: Feature Selection, Medical Image, Sparse feature selection, replica exchange MCMC

1. Introduction

In this research, we compare several feature selection methods for classifying the diffuse lung disease (DLD). The DLD is a kind of inflaming which is spread in the wide are of the lung. In the last stage of the DLD, the disease site lose function of lung and the patient becomes hard to recover, so that early detection of the DLD is desired[1][2] In the DLD sites, there exist large varieties of texture patterns on the CT image. In order to classify the DLD sites, many researchers introduce pattern classification technology into the diagnosis[3][4]. In these methods, the feature extraction is designed manually, however, the combinations of these created features are not so much discussed in this field. Koiwai *et al* introduced the feature selection method using Markov-Chain Monte Carlo (MCMC) method, which is proposed by Nagata *et al.*[5], in this field and showed effective feature combinations. The MCMC method can be regarded as a kind of enumeration method, so that it takes a lot of computational time. Thus, it is better to introduce more effective feature selection method, In this study, we compare some feature selection methods for maximizing the pattern classification performance.

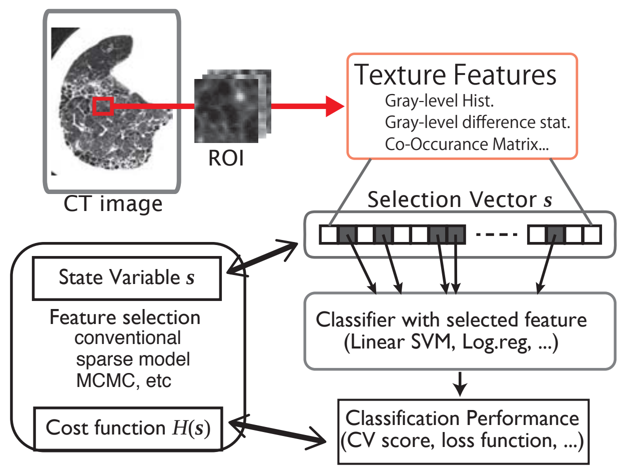

Fig. 1: Schematic diagram of feature selection in this experiments. We extract several texture features from the DLD ROIs. When we select a feature set combination from the set s, we can calculate the classification performance for the score of the selected feature set $H(s)$. The feature selection can formulated as the minimization of the $H(s)$

2. Method

In this section, we explain some feature selection method. In the following, we explain the feature selection framework in this study at first, and then we explain the conventional method. After that, we explain more modern “sparse modeling” based method. At last, we explain about MCMC based

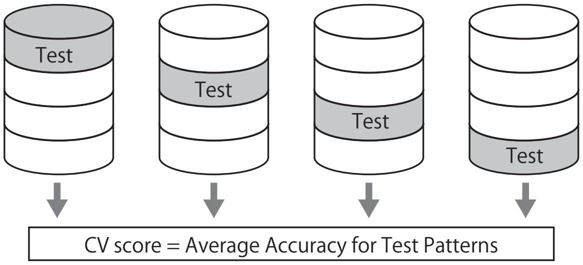

Fig. 2: Measuring method for generalization error using cross validation. Cross validation divides dataset into two parts called "training" and "test" samples. The training set is applied for constructing the classifier, and the test is for evaluation.

method.

2.1 Feature selection framework

Here let us consider the following situation, that is, we have D features in the observation and want to find the most effective features combination for classifying.

The most sure way to find best combination of feature is to search in exhaustive for whole feature combinations. This means the calculation cost for search the best one in the whole combinations is $O(2^D)$, so that this method requires exponential order calculation cost. Fig.1 shows the feature selection framework in this study. We prepare a input data, which are lung disease region images segmented from CT images. From these images, we extract 39 texture features whose details are described in the appendix. After that, we carry out some feature selection algorithm for the features, and evaluate the performance of the classifier for the selected input.

For performance evaluation, we adopt a cross validation (CV) method. Cross validation is a kind of measure for generalization error[6]. Fig. 2 shows a concept for cross validation method. When we divide K sub-dataset, we can choose one subset as a 'test set' and the other as a 'training set'. We train the classifier with training set and evaluate the classification performance with test set. Thus, we can regard the classification performance for the novel input pattern. But it contains some arbitrary selection for the test set, so that we evaluate whole combination of the training set and test set. The cross validation score means the average of the whole combination of the classification performance. This is the concept of the K-fold CV method.

For evaluation task, we use a linear support vector machine (SVM)[7] for conventional methods and MCMC based method. The sparse modeling based methods in this study includes classifier in their models, that is; the cost function for the classifiers are defined as the sum of data fitness term, from which we can derive classifier, and sparse constraint term.

In the following, we explain about the feature selection method in our experiments. Hereafter, we describe the D dimensional input features as vector notation $\boldsymbol{x} = (x_1, \cdots x_D)$. We introduce selection vector $\boldsymbol{s} = (s_1, \cdots s_D)$. Each element takes $s_i \geq 0$ where $s_i > 0$ means we use the ith element and $s_i = 0$ means we do not use.

2.2 Conventional methods

The simple conventional feature selection is based on the mutual information of each feature. Roughly speaking, the feature that has large variety is considered to be informative. So the one conventional method is to evaluate each feature with the χ^2 test. Hereafter, we call this selection method as "univariate selection".

On the other hand, we also evaluate conventional method, which is called "backward" selection. The backward selection is a kind of iterative methods to remove an inefficient feature in the feature set in each iterations. When classification performance does not not significantly change with subtracting the feature, the selection iteration is stopped. The backward selection decrease the features in its selection process. So, usually, the initial feature set starts with full feature set $\boldsymbol{s} = \boldsymbol{1} = (1, 1, \cdots, 1)$. Then, we choose the most ignore-able feature in the Dth elements. When the ith element is the one, we mark it as the $s_i = 0$. If removing the ith element does not improve the classification performance, we omit the removing the ith element and finish the algorithm. Hereafter, we call this selection method as "recursive selection".

2.3 Sparse modeling based method

The conventional method is an algorithm for single selection choosing process. The forward/backward selections treat only one feature at a time in the iterations. Miss-selection might be occurred when correlated features are given in the features. So that it is better to treat plural features simultaneously at a time. The "sparse" modeling based method, carry out the simultaneous optimization for data-fitting and some constraint regularization. In this study, we apply both the cross-entropy and the hinge-loss as the data-fitting part. The former derives the logistic regression, and the latter derives the SVM. The constraint part in this study, we adopt the ℓ_1 constraint.

Here, we consider the one-versus-rest classification problem with logistic regression. The index p shows the pattern index, and $\sigma(u)$ denotes the logistic function $\sigma(u) = 1/(1 + \exp(-u))$. We also consider the training data pair

$\{\boldsymbol{x}_p, t_p\}$ where $\boldsymbol{x}_p$ means pth input datum, and t_p shows the annotation label where $t_p = 1$ means the pth input belongs to the desired class, and $t_p = 0$ means the other case. When we denote the $y_p = \sigma(\boldsymbol{s}^{\mathrm{T}}\boldsymbol{x}_p)$ as the probability the $\boldsymbol{x}_p$ belongs to the desired class, the cost function of cross-entropy can be denoted as

$$\tilde{H}_{\text{cross}}(\boldsymbol{s}) = -\sum_p t_p \log y_p + (1 - t_p)\log(1 - y_p). \quad (1)$$

In order to carry out sparse selection, we introduce ℓ_1 constraint, so that the total cost function becomes

$$H_{\text{cross}}(\boldsymbol{s}) = \lambda \tilde{H}_{\text{cross}}(\boldsymbol{s}) + \sum_i |s_i|, \quad (2)$$

where λ shows the hyper-parameter to control the balance between the data-loss and constraint loss terms. Then we carry out the minimization of $H(\boldsymbol{s})$, we can obtain the sparse representation which can be regard as a feature selected solution. When we optimize the parameter, and obtain a solution $\tilde{s}$ for the cost function 2, we obtain a ℓ_1 constraint logistic regression classifier $y = \sigma(\tilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{x})$, which means the probability of the input $\boldsymbol{x}$ belongs to the desired class. Hereafter, we describe this as "L1 sparse logistic regression" (L1-SLR) classifier.

In the same way, we can consider the hinge-loss function. the cost function of hinge-loss can be denoted as

$$\tilde{H}_{\text{hinge}}(\boldsymbol{s}) = \sum_p \max\{1 - t_p y_p, 0\}, \quad (3)$$

where t_p shows the annotation label where $t_p = 1$ means the pth input belongs to the desired class and the $t_p = -1$ means the other case. The function y_p means the output value of the linear decision function $y = \boldsymbol{s}^{\mathrm{T}}\boldsymbol{x} + b$ for the input $\boldsymbol{x} = \boldsymbol{x}_p$ Here, we also derive sparse solution, we introduce ℓ_1 sparse prior for the loss function. Then we optimize the following loss function:

$$H_{\text{hinge}}(\boldsymbol{s}) = \lambda \tilde{H}_{\text{hinge}}(\boldsymbol{s}) + \sum_i |s_i|. \quad (4)$$

The standard SVM loss function is defined as the second term with ℓ_2 prior $\|\boldsymbol{s}\|^2$ regularization, however, in eq.(4), we substitute the regularization term with ℓ_1 form for pruning. In order to obtain the solution, we also optimize the parameter and obtain a solution $\{\tilde{Vecs}, \tilde{b}\}$ for the cost function 4, and obtain a ℓ_1 constraint support vector classifier $y = \tilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{x} + \tilde{b}$, where y means the score of the input $\boldsymbol{x}$. The score $y > 0$ means the input $\boldsymbol{x}$ belongs to the desired class. Hereafter, we describe this as "L1-SVM" classifier.

2.4 MCMC based method

We introduce MCMC based method in the manner with Nagata *et al.* Here, we introduce linear SVM as the classifier

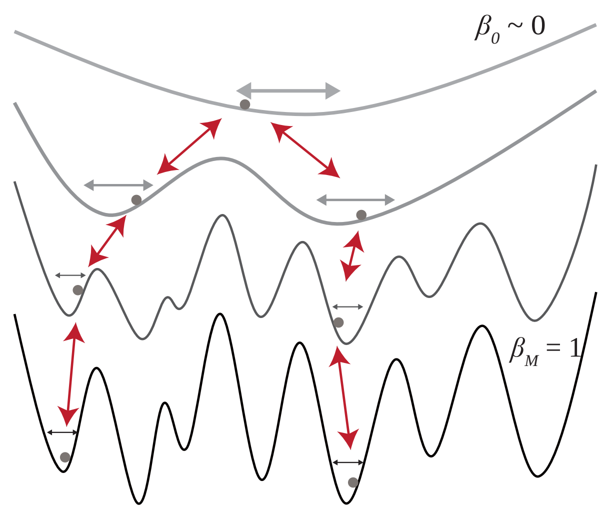

Fig. 3: Landscape of the the replica exchange MCMC method. The horizontal axis shows the state space, and the vertical shows the cost function $H_m(\boldsymbol{s})$. Preparing parallelized Monte Carlo sampling system with different inverse temperature systems, the state vectors move into another local minima easily via low β system.

for this method. The linear SVM classify the input with decision boundary $y(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b = 0$[7].

The purpose of the MCMC based method is to enumerate the solution candidate of $\boldsymbol{s}$ with stochastic sampling. In the manner of the Nagata's method, we introduce the cost function $H(\boldsymbol{s})$ as the CV score. Using optimization method for minimization $H(\boldsymbol{s})$, we can only find a candidate of $\boldsymbol{s}$. On the other hands, the MCMC based method can propose plural solutions. From this point of view, MCMC based method is not a bad strategy for obtaining solution. The procedure for the MCMC is summarized as following:

1) Select one site $s_i^{(t)}$ in the state vector $\boldsymbol{s}^{(t)}$ where t means the time index.
2) Prepare a candidate vector $\boldsymbol{s}^*$ in which only $s_i^{(t)}$ is inverted from the vector $\boldsymbol{s}^{(t)}$.
3) Calculate the costs $H(\boldsymbol{s}^{(t)})$ and $H(\boldsymbol{s}^*)$ and evaluate the probability

$$r = \min\left(1, \frac{\exp(-H(\boldsymbol{s}^*))}{\exp(-H(\boldsymbol{s}^{(t)}))}\right). \quad (5)$$

4) Generate an unit random value $u \in [0, 1]$, and compare u with the r. If $u < r$ then accept the state $\boldsymbol{s}^*$ as a new state $\boldsymbol{s}^{(t+1)}$, and the other case the state is hold as $\boldsymbol{s}^{(t+1)} = \boldsymbol{s}^{(t)}$
5) Goto the 1st step while t satisfies the iteration limit

This method is known as Metropolis-Hasting (MH)method[8], and hereinafter we call this successive procedures as Monte Carlo step(MCS). Using the MCMC method, we can obtain samples $\{s^{(t)}\}$ which obeys the probability $p(s) \propto \exp(-H(s))$.

The MH method is a strong method for enumerating, however, it requires long calculation time to sample from wide spreading multiple peak distribution. For transition from a peak to another, there exists low probability region in any transition paths. The driving force of the MCMC depends on the odds ratio of the pre- and post- state in eq.(5). So that, too much low probability region prohibits desirable transition.

The replica exchange MCMC method is to overcome the transition problem[9][5]. We introduce temperature parameter $T > 0$ and its inverse $\beta = 1/T$. Considering the probability with inverse temperature β of probability, we can re-define the probability with weight by inverse temperature $p(s) \propto \exp(-\beta H(s))$. The temperature $\beta = 1$ means our original cost function. When β becomes small, the efficacy of the cost function $H(s)$ also becomes small. So the landscape of the weighted cost function $\beta H(s)$ becomes smooth. Fig.3 shows the concept of the replica exchange method. We prepare L parallel replicated MCMC system, and we run each MCMC with different temperature T_l. After several MCS, we exchange several replica states. As the result, we can obtain sample from wide spreading multiple peak distribution via low temperature Markov-chain transitions. The procedure for the ExMCMC is summarized as following:

1) Prepare M replicated systems, and assign appropriate inverse temperature $0 < \beta_0 < \beta_1 < \cdots < \beta_{M-1} = 1$. Denoting each system status variable as s_m where m means the index of system.
2) Carrying out several MCSs under the probability of $p(s_m) \propto \exp(-\beta_m H(s_m))$ for mth system. Now, we describe the exchange timing as τ.
3) Select one temperature site denoted as j.
4) Calculate transition probability $W({s^{(\tau)}}_j, {s^{(\tau)}}_{j+1})$ as

$$W({s^{(\tau)}}_j, {s^{(\tau)}}_{j+1}) = \min(1, \exp(-\Delta)) \quad (6)$$

$$\Delta = (\beta_j - \beta_{j+1})(H({s^{(\tau)}}_j) - H({s^{(\tau)}}_{j+1})) \quad (7)$$

5) Generate a unit random variable $u' \in [0, 1]$, and compare it with $W(s_j, s_{j+1})$.
6) If $u' < W({s^{(\tau)}}_j, {s^{(\tau)}}_{j+1})$, exchange the states ${s^{(\tau)}}_j$ and ${s^{(\tau)}}_{j+1}$.
7) Goto 2 for several MCSs.

So that, this parallel MCMC mechanism work as a outer loop of the each MH method. Applying the replica exchange MCMC method, escape from the the local minima is just easier rather than that of the single MCMC method. Hereafter, we call this obtained classifiers as exchange MCMC (ExMCMC) classifiers.

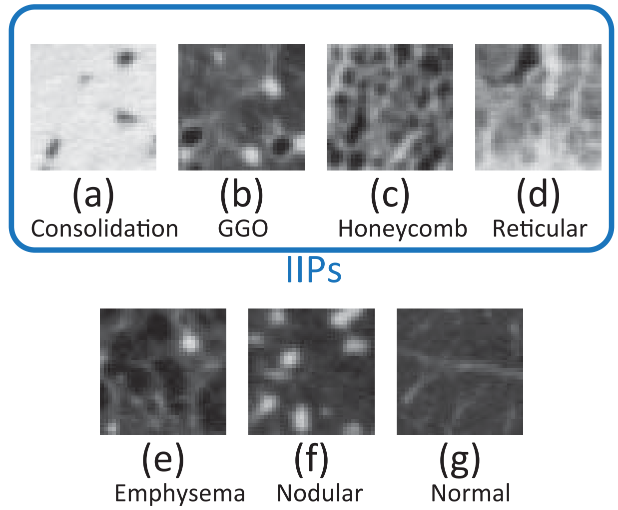

Fig. 4: Typical ROI samples for each class. Classes from (a) to (d) are involved in the IIPs. Classes (e) and (f) is another disease. The class (f) means the normal class.

3. Experiments

3.1 Materials

In this research, we prepare 360 labeled images. Each class has following number of images: Consolidation(CON):38, Ground-Grass-Opacity (GGO):76, Honeycomb(HCM):49, Reticular(RET):37, Emphysema(EMP):54, Nodular(NOD):48, and Normal(NOR):58 cases. We assume the 32×32 [pixels] ROIs, and each ROI is segmented under the direction of a physician, and diagnosed by 3 physicians.

The acquisition parameters of those HRCT images are as follows: Each images are obtained from Toshiba "Aquilion 16" imaging device. Each slice image consists of 512×512 pixels, and pixel size corresponds to $0.546 \sim 0.826$ [mm], slice thickness are 1 [mm]. The origin of these image data is provided Tokushima University Hospital. Fig.4 shows a typical image example of each disease in HRCT image. The left shows an overview of the axial HRCT images of lungs including lesion, and the right shows segmented images of typical examples of lesion from the left image collections. The consolidation (CON) and ground-grass opacity (GGO) patterns are often appeared with the cryptogenic organizing pneumonia diseases (COPD). The GGO pattern is also often appeared in the non-specific interstitial pneumonia (NSIP). The reticular (RET) pattern which sometimes includes GGO patterns is also appeared in the NSIP. The honeycomb (HCM) pattern has more rough mesh structure rather than that of the RET, and it appeared in idiopathic pulmonary fibrosis (IPF) or usual interstitial pneumonia (UIP).

We introduce several texture representations proposed by Sugata *et al.* for features[10][1]. From the input HRCT ROI image, we calculate gray-level histogram (GLH), gray-level difference (GLD) statistics, the co-occurrence matrix (COM), run length matrix (RLM), and Fourier power spectrum (FPS). From each of GLH, GLD, FPS of radial direction denoted as FPS(r), and FPS of angle direction denoted as FPS(θ), we extract mean, contrast, variance, skewness, kurtosis, energy and entropy. From the co-occurrence matrix, we extract energy, contrast, correlation, variance, entropy. From the run length matrix, we extract short/long run emphasizes, gray level no-uniformity, run length no-uniformity, and run percentage. At last we obtain 39 texture features from one ROI image.

3.2 Computational environment

For the control experiment, we prepare a standard SVM classifier using whole features. The SVM parameters for soft-margin control is determined with cross validation of $k = 5$. The number of categories is 7, so that we adopt 'one-versus-rest' (OVR) method to apply the multi-class classification. The OVR method construct a class specific classier for one class, so that, the one classifier identify whether the input is belong to the class or not.

In the univariate selection and the recursive selection, we choose the feature set that maximizes the CV score. We use *scikit-learn* package for these methods[11].

In the sparse modeling based selections that are L1-SLR and L1-SVM, at first, we choose the hyper-parameters λ from $\{0.1, 0.2, \cdots, 1.7\}$ and fix them as 1.0 experimentally with the CV-score. After that, we carry out minimization the loss functions (2) and (4). For these methods, we also use *scikit-learn* package for these methods.

In the MCMC based, we carry out the selection with the manner of Koiwai *et al*[12]. We prepare $M = 7$ temperature replica systems, and iterate $T_{\max} = 20,000$ times. For the evaluation of the MCMC state, we apply 10-fold CV score as the cost function $H(\boldsymbol{s})$. We apply MATLAB for this methods.

4. Results

Table 1 shows the CV score for each class with selection methods. The first row shows the results of control experiments, that is the result with whole features. The performance for the CON, GGO, HCM, RET, EMP classes shows the over 0.9, however, the one for the NOD is not good result. The average of these scores is 0.86. The conventional feature selections, which are univariate and recursive selections improves the score for NOD class. However, in the univariate selection, the ones for GGO, HCM, RET, and NOR classes becomes slightly worse. The recursive selection shows better results rather than the control and univariate selection, however, the result with NOD class is just worse than the univariate selection. The average score of univariate feature selection is 0.90, and the one of recursive selection is 0.89. Both sparse modeling based methods shows improvement in the NOD class performance. The average score of L1-SLR is 0.90 and the one of L1-SVM is 0.93. The MCMC based method also shows the good performance. Especially, the only this method shows the over 0.9 score for NOD class performance. In the GGO and EMP classification, the classification score L1-SVM and L1-SLR is just superior than that of the MCMC based method respectively. This means the MCMC method fails to search the solution.

Table 1: Classification performance of the CV score for feature selections for each class

	CON	GGO	HCM	RET	EMP	NOD	NOR
Whole	1.000	0.915	0.972	0.958	0.915	0.380	0.888
Univar.	1.000	0.817	0.958	0.901	0.901	0.873	0.845
Recur.	1.000	0.930	0.972	0.958	0.986	0.521	0.888
L1-SLR	1.000	0.915	0.972	0.958	0.972	0.648	0.859
L1-SVM	1.000	0.958	0.972	0.944	0.915	0.775	0.915
MCMC	1.000	0.939	0.989	0.980	0.967	0.925	0.981

Fig. 5 shows the selected features for each class. In each figure, each row shows the selected feature for each method. The horizontal axis shows the texture features index. The colored location shows the selected feature and white location shows the non-selected feature. The depth of colors in L1-SLR, L1-SVM, and MCMC based method shows the strength of the linear classifier. Comparing with the MCMC based methods, other feature selection method is just too much sparse. In the L1-SLR and L1-SVM methods, this sparsity can control with the strength of the parameter λ in eq(2) and eq(4). However, optimization of the λ parameters requires more calculation cost. In the GGO class, comparing with the L1-SVM method, the MCMC based method looks similar selection, but it has much more selected features in the GLD and GLH features. In the EMP class, comparing with the L1-SLR method, the MCMC based method also takes FPS(θ) features.

5. Conclusion & Discussion

In this research, we compare feature selection methods of conventional method, sparse modeling based method and MCMC based method for DLD classification task. The conventional methods, which are based on the univariate

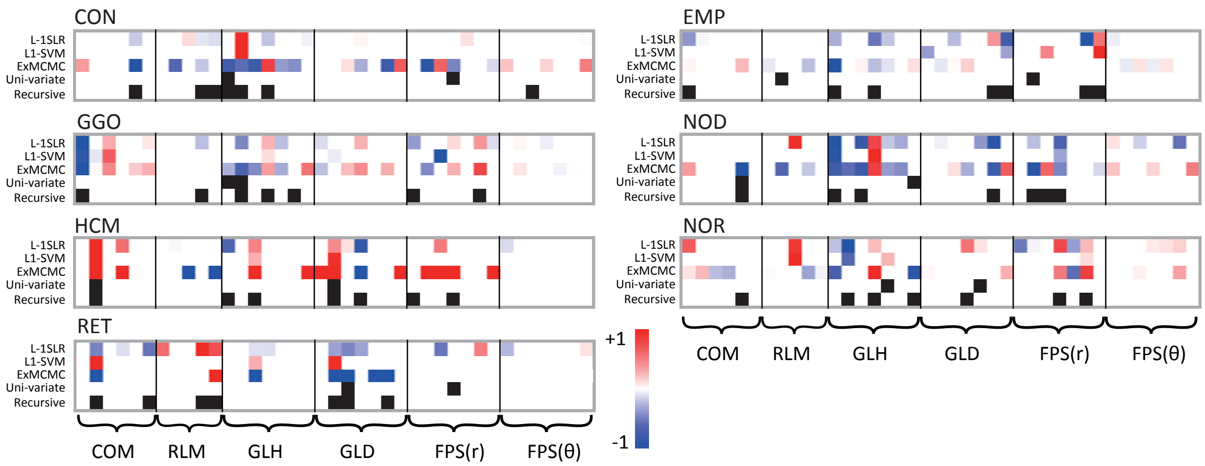

Fig. 5: In each figure, each row shows the results of selected features for each method. The horizontal axis shows the index of extracted features. The color depth in the MCMC, L1-SLR, and L1-SVM raws show the strength of the weights for linear classifier.

feature selection and recursive feature elimination methods, does not show the good performances for this task. These methods treat feature selection problem as a single feature importance in iteration steps. Hence, these methods might not be able to handle the correlated features well. The sparse modeling based methods, which are L1-SLR and L1-SVM methods, shows better results in the meaning of the classification score rather than those of the conventional methods. Comparing with the MCMC based method, these methods show better performance in the particular classes. The MCMC based methods shows the best performance in the average score and almost all the classes. However, in the several classes, the selected features are just too much rather than the best solution in these experiment.

In the future work, we would like to improve feature selection method. The MCMC based method is just too much calculation cost, so that we would like to combine the sparse modeling based feature selection method with the one with MCMC based method.

Acknowledgment

We thank Professor Junji Ueno, Tokushima University. He provided several advice for this study as well as a set of high resolution HRCT image of IIPs. This work is supported by Grant-in-Aids for Scientific Research (C) 16K00328, and Innovative Areas 16H01542, MEXT, Japan.

References

[1] Y. Sugata, S. Kido, and H. Shouno, “Comparison of two-dimensional with three-dimensional analyses for diffuse lung diseases from thoracic ct images,” *Medical Imaging and Information Sciences*, vol. 25, no. 3, pp. 43–47, 2008. [Online]. Available: http://ci.nii.ac.jp/naid/130000097652/en/

[2] M. Wada, H. Shouno, and S. Kido, “An idiopathic interstitial pneumonia classification for ct image by use of a semi-supervised learning,” in *Intl. Forum on Medical Imaging in Asia (IFMIA)*, November 2012, pp. P1–34.

[3] R. Xu, Y. Hirano, R. Tachibana, and S. Kido, “Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach,” in *MICCAI*, vol. 14. Springer-Verlag Berlin Heidelberg, 2011, pp. 183–190.

[4] H. Shouno and S. Kido, “Semi-supervised based learning for Idiopathic Interstitial Pneumonia on High Resolution CT images,” in *In Proc. PDPTA*, Jul. 2015, pp. 270–275.

[5] K. Nagata, J. Kitazono, S.-i. Nakajima, S. Eifuku, R. Tamura, and M. Okada, “An exhaustive search and stability of sparse estimation for feature selection problem,” *IPSJ Transactions on Mathematical Modeling and Its Applications*, vol. 8, no. 2, pp. 23–30, 2015.

[6] M. Stone, “Cross-validation: A review.” *Math.Operations.Stat.Ser.Stat*, vol. 9, no. 1, pp. 127–139, 1978.

[7] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.

[8] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines,” *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[9] H. Koji and N. Koji, “Exchange monte carlo method and application to spin glass simulations,” *Journal of the Physical Society of Japan*, vol. 65, no. 6, pp. 1604–1608, 1996. [Online]. Available: http://dx.doi.org/10.1143/JPSJ.65.1604

[10] R. Uppaluri, E. Heitmman, M. Sonka, P. Hartley, G. Hunninghake, and G. Mclennan, “Computer recognition of regional lung disease patterns.” *American Journal of Respiratory and Critical Care Medicine*, vol. 160, no. 2, pp. 648–654, 1999.

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[12] M. Koiwai, N. Iida, H. Shouno, and S. Kido, “Feature selection for diffuse lung disease using exchange markov chain monte-carlo method,” in *PDPTA*. PDPTA 2016, 7 2016, pp. 381–386.