Classification of drug tablets using hyperspectral imaging and wavelength selection with a GAWLS method modified for classification

Hiromasa Kaneko and Kimito Funatsu

Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo,
Bunkyo-ku, Tokyo 113-8656, Japan

Corresponding author: Kimito Funatsu

Tel: +81-3-5841-7751

Fax: +81-3-5841-7771

E-mail: funatsu@chemsys.t.u-tokyo.ac.jp

Abstract

Right drug tablets must be brought to the right places. We apply hyperspectral imaging, which can measure infrared spectra at many points on a two-dimensional plane, to classify tablets correctly. The k-nearest neighbor algorithm (kNN) is employed to classify tablets using a database including their spectra and true classes. Although classification accuracy is not 100%, we can correctly classify tablets overall, since spectra at many points are measured with spectroscopy and misclassification at some points does not have much influence on the final tablet classification result. In addition, we propose a wavelength selection method for classification. Genetic algorithm-based wavelength selection is applied to classification and combined with kNN, and thus, not wavelengths but wavelength-regions can be selected in classification problems. Through a case study, we confirmed that the proposed method could classify three kinds of tablets correctly and select appropriate wavelength-regions.

Keywords: Hyperspectral imaging; Classification; Tablets; k-nearest neighbor; Wavelength selection; Variable selection.

1 Introduction

When many kinds of drug tablets are handled, it is important to bring the right tablets to the right places. Misclassification, which means that drug tablets are wrongly classified, is a crucial problem. In addition, false tablets, which are dangerous for health, are sometimes marketed illegally. Three types of false drug tablets exist; those that do not contain any active pharmaceutical ingredient (API), the ones containing API that are not marked in the packing and drug tablets that contain the marked API, but are produced by a different manufacturer (World Health Assembly, 2002; Rodionova et al., 2005). Before we acquire them, they must be checked accurately, non-destructively and rapidly. Since false drug tablets have such a high quality nowadays, it is difficult to detect them.

Process analytical technology (Roggo et al., 2007; Hansuld and Briens, 2014) is an important technique for monitoring, developing, controlling and designing critical product quality in the pharmaceutical industry. Near infrared (NIR) spectroscopy (Muteki et al., 2013), Raman spectroscopy (Simone et al., 2014) and so on have been focused to monitor product quality non-destructively in real time.

In this study, we focus on hyperspectral imaging, which provides spatial and spectral information simultaneously. For samples on a scene, a hyperspectral image sensor outputs the spectrum for each pixel in the image. By using near-infrared spectroscopy as a spectroscopic method, measurements of samples can be performed accurately, non-destructively and rapidly. Hyperspectral imaging has been applied to prediction of protein and fat content in cheeses (Burger and Geladi, 2006), estimation of fruit yield in citrus (Ye et al., 2006), mapping piroxicam polymorphs (Rocha, 2011), assessment of regional leaf area index (Canisius and Fernandes, 2012), food authentication (Ottavian et al., 2012), discrimination between polylactic acid and polyethylene terephthalate (Ulrici et al., 2013) and monitoring of polymorphic transformation (Terra and Poppi, 2012). Pre-treatment (Esquerre et al. 2014) and spatial resolution (Offroy, 2012) have been discussed as well. Gowen et al. provided a review paper of hyperspectral imaging as a process analytical tool for food quality and safety control (Gowen, 2007).

Alvarez-Jubete et al. (2013) constructed regression models between API concentrations and NIR spectra in pixels for hyperspectral imaging. Vajna et al. (2011) proposed to estimate concentrations in the extruded pharmaceutical by using Raman chemical imaging and multivariate curve resolution-alternating least squares (MCR-ALS). Carneiro et al. (2014) estimated different chemical compositions between

crystals and cream using infrared imaging spectroscopy and MCR–ALS. Edelman et al. (2013) applied hyperspectral imaging to color classification of ecstasy tablets. They succeeded in determining differently colored tablets and estimating concentration of a colorant. Rodionova et al. classified a legitimate drug tablet and a counterfeit drug using hyperspectral imaging (Rodionova et al., 2005). To construct the classification model, a soft independent class modeling analogy, which is a supervised pattern recognition method, was applied to NIR spectra. However, they do not consider drug tablets whose true classes do not exist in training data. When spectra belonging to these tablets are input into a classification model, they are classified into one of the classes existing in training data, which is not true.

In this paper, our objective is to classify drug tablets including those whose true classes do not exist in training data, by using hyperspectral imaging. First, for several tablets whose drug classes are known, spectra are measured with a hyperspectral image sensor. Then, a statistical model is constructed between spectra and the class labels of the tablets. The classes of new tablets can be estimated by inputting their spectra measured with a hyperspectral image sensor into the constructed model. Although the classification accuracy of a model is not 100%, we can correctly classify tablets overall, since spectra at many points are measured with spectroscopy and misclassification at some points does not have much influence on the final tablet classification result.

There are several classification methods such as support vector machine (SVM) (Bishop, 2006) and partial least squares-discriminant analysis (Toher, 2007), but we employ k-nearest neighbor algorithm (kNN) (Kaneko and Funatsu, 2013a), which is a simple method, easy to be applied to multiclass classification, and can handle nonlinear relationships between spectra and class labels. By using kNN model, we can classify tablets only from spectra measured with a hyperspectral image. kNN can provide simpler models than SVM, which is designed for binary classification problems originally, and it can be applied to multi-class classification (Crammer and Singer, 2001). The performance of kNN is sufficiently high, as shown in “3 Results and Discussion”. In addition, unknown classes, which do not exist in training data, can be handled using the distance threshold in kNN. When spectra of drug tablets, whose true classes do not exist in training data, are input into a classification model and exceed the threshold, they are classified as a class that does not exist in training data.

Wavelength selection of spectra is effective for the improvement of statistical models, including kNN models, and the cost reduction of spectral measurement. Many wavelength selection methods have been developed (Arakawa et al., 2011; Kim et al. 2011). Specifically, genetic algorithm-based wavelength selection (GAWLS), which has

been used only for regression analysis, can select multiple wavelength-regions (Arakawa et al., 2011). It was confirmed that reasonable wavelength-regions could be selected and appropriate regression models could be constructed using the selected wavelength-regions, where GAWLS is superior over some other wavelength selection methods. GAWLS can be applied to time series data analysis (Kaneko and Funatsu, 2012) and nonlinear systems (Kaneko and Funatsu, 2013b).

We therefore apply GAWLS to classification problems in order to improve predictive ability of classification models and select appropriate wavelengths. Accuracy rate using cross-validation (CV) in kNN modeling is employed as a fitness function in a genetic algorithm (GA), where we will be able to select combinations of wavelength-regions with which predictive classification models can be constructed.

The effectiveness of our proposed method is demonstrated through a case study using real drug tablets whose spectra are measured with a hyperspectral image sensor. We show that drug tablets can be classified accurately and reasonable wavelength-regions can be selected. New classes that do not exist in training data can be handled by considering the distance threshold in kNN.

2 Method

2.1 kNN

kNN is a well-known classification method. To classify a query sample, we simply determine the class to which most k known samples are the closest to the query sample belong. We use the Euclidean distance between two samples as a distance measure.

We can handle samples whose classes do not exist in training data by setting a threshold for some distance in kNN as applicability domain (Kaneko and Funatsu, 2013a; Escobar et al. 2014; Kaneko and Funatsu, 2014), which is a data domain where a constructed model has highly predictive performance. The average of distances, the maximum distance, or the minimum distance from k known data that are the closest to the query data can be used, for example. In this paper, we employ the average of the distances from k known data and set the threshold using 5-fold cross-validation (CV). The threshold is the 99.7% cross-validated value of the average distances.

2.2 GAWLS (Arakawa et al., 2011)

GAWLS is one of the methods that is used to select combinations of important wavelengths (variables) from explanatory variables (X) using regions as a unit of measurement. A genetic algorithm (GA) (Leardi, 2001) is applied to select wavelengths.

GA is an optimization method that is used in biology to model principles of natural evolution. Species having a high level of fitness under certain environmental conditions can prevail in the next generation, and the best species may be reproduced by crossover together with random chromosomes mutation in those species that survive. The solution space around superior individuals is preferentially explored, which leads to discovering a solution that is close to the optimum.

In GAWLS, two actual genes of a chromosome represent one region of wavelengths. Figure 1 shows the coding method for GAWLS. Hence, GAWLS can select important wavelengths using regions as a unit of measurement. In Figure 1, the number of selected wavelengths is 1265 through 1283 and 2057 through 2069, and the number of windows (NW) is 2. Partial least squares (PLS) model (Wold et al. 2001), which is a regression model, is constructed then with a set of selected wavelengths. The r_{CV}^2 -value, which is the determinant coefficient calculated using a CV method is used as a fitness value for each chromosome. In this way, a set of an optimum wavelength-regions, with which highly predictive model can be constructed, is obtained.

The number of chromosomes, the number of generations, the maximum width of window and NW must be set beforehand. The minimum width is zero.

2.3 GAWLS for classification

We propose to apply GAWLS to classification problems. Figure 2 shows the basic concept of the proposed method. To calculate a fitness value for each chromosome, CV based on kNN is performed with only the selected wavelength as input variables. The fitness value is the cross-validated accuracy rate (AR) given as follows:

$$AR = \frac{N_{\text{correct}}}{N_{\text{all}}} \quad (1)$$

Here, N_{all} denotes the number of all samples and N_{correct} represents the number of samples for which drugs are correctly classified in CV. The proposed method can select wavelength-regions where high predictive classification models can be constructed.

We used the Genetic Algorithm Optimizing Toolbox (Houck et al., 1991) for the genetic algorithm calculation.

2.4 Data

A set of three commercially available tablets (Tablet A, B and C) are prepared.

The types of API are different; acetylsalicylic acid in tablet A, berberine tannate in tablet B and anhydrous caffeine in tablet C. ImSpector (<http://www.jfe-tec.co.jp/product/hikari/imspector.html>) was used as a hyperspectral image sensor. The size of hyperspectral images is 320×253 pixels. Each pixel has one spectrum, i.e. one image has 80960 (=320×253) spectral data. Spatial resolution is about 0.28 mm, spectral range is from 932.0 to 2390.8 nm, and spectral resolution is about 5.8 nm. Before sample measurement, background spectra of a sample stage were measured, obscuring light. The background spectra are subtracted from measured spectra of samples and a sample stage. For each pixel and at each wavelength, the reflectance value (R) is calculated from the raw intensity value (I) using the dark current (D) and the intensity of the light source (L) as follows:

$$R = \frac{I - D}{L - D}. \quad (2)$$

We used two hyperspectral images or batches; one set is the training data, where the classes of the tablets are assumed to be known. The other set is the test data, which is used as an external validation set that is separate from the training data set. Each data set includes different A, B and C tablets (the number of all tablets is six). The number of the training data is 2180 and that of the test data is 2119. The number of classes is four, including background data, which means that there is no tablet on a stage. The number of training and test data is referred to the number of pixels. 249 for tablet A, 293 for tablet B, 406 for tablet C and 1232 for background in training data. 247 for tablet A, 274 for tablet B, 409 for tablet C and 1169 for background in test data. The number of original input variables is 253.

3 Results and Discussion

3.1 Classification

As a preprocessing method, Savitzky–Golay (SG) filter (Savitzky and Golay, 1964) was employed. SG filter can differentiate and smooth spectra simultaneously, where the details of the SG method can be found in the literature (Luo, 2005; Yoshimura and Takayanagi, 2012). Width is 21, polynomial order is 2 and derivative order is 2. k in kNN is 10. These parameter-values were optimized with 5-fold CV using only training data. Then, spectra of training data were auto-scaled and those results were used for test data scaling as well.

We assumed that the classes of training data were known already and these data

were stored in the database. Then, test data classes were predicted. Table 1 shows the prediction results. The AR of the test data exceeds 0.99, reflecting the highly predictive ability of the kNN model. Someone may worry that the AR-value is lower than 1. However, as shown in Figure 3 including classification results of a test hyperspectral image, we could completely distinguish tablets by using hyperspectral imaging, even when the AR-value is not 1. It was confirmed that types of tablets could be distinguished correctly by combining hyperspectral imaging and kNN classification.

However, new tablets whose classes are different from the classes in training data may be input into our classification system. Therefore, we assumed that one of the three tablet-classes is unknown for testing and this was repeated three times so that data of each class are used once as the unknown class. By setting the threshold for the kNN distance (see Section 2.1), we can judge whether a new query sample can be classified or not, *i.e.*, whether a new query is an existing class or an unknown class. When the kNN distance of a new query exceeds the threshold, the query is assigned as an unknown class.

When the unknown class is changed three times, receiver operating characteristic (ROC) curves are shown in Figure 4. The ROC curve is a plot of true positive rate against false positive rate. Table 1 shows AUC values, which are the area under ROC curves. As shown in Figure 4, ROC curves are idealized ones and AUC values are sufficiently high. We can say that the proposed method is able to correctly discriminate both samples that are known classes and samples that are unknown classes.

3.2 Wavelength selection

Then, we tried to select appropriate wavelength-regions with which highly predictive classification models can be constructed. The number of chromosomes was set to 300 and the number of generations was set to 500. The probability of crossover, the probability of mutation and the rate of selected chromosomes were set as default values, 0.6, 0.05 and 0.3, respectively. The maximum width of window was set as 30 and NW was changed to 1, 2, 3, 4 and 5. We performed GAWLS for classification 10 times, *i.e.*, 10 runs for each NW.

The prediction results of test data for each NW are shown in Table 1. For each NW, AR showed high values, which means high predictive classification ability. Although the maximum number of NW, 5, is an arbitrary number, larger NW numbers are not required because AR values are sufficiently large and do not change so much. Even when NW is only 1, AR exceeds 0.99 and is on the same level as that of kNN before wavelength selection. Since the prediction results are almost the same as those in

Figure 3 when the selected wavelengths are only one region, all types of tablets can be distinguished correctly by using both hyperspectral imaging and kNN. Then, other three data sets that include different A, B and C tablets were prepared, were the proposed GAWLS-kNN models predicted their classes. For all models and all data sets, AR values exceeded 0.99. We confirmed that appropriate wavelength-region(s) could be selected, leading to cost reduction of hyperspectral image sensor and realistic tablet discrimination models. In our case study, we only used different tablets including different API types, achieving accurate classification of tablets. The performance of the proposed method should be evaluated considering tablets with the same type of API and different amounts of API.

Figure 5 shows the selected wavelength regions in GAWLS when NW is 1, 2 and 3 and Figure 6 shows frequency of selection made for all five NW values. The uniform regions are selected for each NW although selected regions are sometimes different from the majority. One reason for such difference in selected wavelength regions would be that correlation structure in selected wavelengths varies due to the change of the number of selected wavelength regions. Intensity around 1360, 1690 and 2240 is attributed to a methyl group and intensity around 2080 and 2380 is attributed to a hydroxy group. Because of that, presence or absence of a methyl group and that of a hydroxyl group could be judged by selecting wavelength-regions around 1360, 1690 or 2240, and 2080 or 2380, leading to accurate classification results. When NW is 1, the selected wavelength-region would be attributed only to a methyl group. Correct classification could be achieved with only the selected wavelength-region, because the amount of methyl group is different in tablets A, B and C and, accordingly, the intensity is different. Spectroscopists would be able to discover previously unknown, but important wavelength regions, since all models constructed using the selected regions indicated high predictive ability.

4 Conclusions

In this paper, a new classification method based on hyperspectral imaging and kNN is proposed to distinguish tablet classes from only the information on those classes in the database and spectrum of objective tablets. The case study using a real data set showed that the proposed method could completely classify not only the classes in training data but also the classes where no training data existed.

In addition, by using the proposed GAWLS method for classification, i.e. GAWLS-kNN, small, but reasonable, wavelength-regions could be selected appropriately. The measurement cost of a hyperspectral image sensor can be reduced

dramatically by using the proposed method.

In PAT fields, it is desired that all right tablets are brought to the right places and the proposed method is sure to contribute it and can detect anomalous and dangerous drugs beforehand.

Acknowledgements

The authors acknowledge the support of the Core Research for Evolutionary Science and Technology (CREST) project “Development of a knowledge-generating platform driven by big data in drug discovery through production processes” of the Japan Science and Technology Agency (JST). The authors acknowledge the support of JFE Techno-Research Corporation. The authors are grateful for the precious support of Ms. Hitomi Onodera in data analysis.

References

- [1] World Health Assembly, Counterfeit Drugs: Threat to Public Health, vol. 55, World Health Assembly, Geneva, 2002.
- [2] Rodionova, O.Y., Houmøller, L.P., Pomerantsev, A.L., Geladi, P., Burger, J., Dorofeyev, V.L., Arzamastsev, A.P., 2005. NIR spectrometry for counterfeit drug detection: A feasibility study. *Anal. Chim. Acta.* 549 (2005) 151-158.
- [3] Y. Roggo, P. Chaluz, L. Maurer, C.L. Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharmaceut. Biomed.* 44, 683–700.
- [4] Hansuld, E.M., Briens, L., 2014. A review of monitoring methods for pharmaceutical wet granulation. *Int. J. Pharm.* 472, 192–201.
- [5] Muteki, K., Blackwood, D.O., Maranzano, B., Zhou, Y., Liu, Y.A., Leeman, K.R., Reid, G.L., 2013. Mixture component prediction using iterative optimization technology (calibration-free/minimum approach). *Ind. Eng. Chem. Res.* 52, 12258–12268.
- [6] Simone, E., Saleemi, A.N., Nagy, Z.K., 2014. Application of quantitative Raman spectroscopy for the monitoring of polymorphic transformation in crystallization processes using a good calibration practice procedure. *Chem. Eng. Res. Des.* 92, 594–611.
- [7] Burger, J., Geladi, P., 2006. Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst* 131, 1152–1160.
- [8] Ye, X., Sakai, K., Garciano, L.O., Asada, S., Sasao, A., 2006. Estimation of citrus

yield from airborne hyperspectral images using a neural network model. *Chemom. Intell. Lab. Syst.* 198, 426–432.

- [9] Rocha, W.F.C., Sabin, G.P., Março, P.H., Poppi, R.J., 2011. Quantitative analysis of piroxicam polymorphs pharmaceutical mixtures by hyperspectral imaging and chemometrics. *Chemom. Intell. Lab. Syst.* 106, 198–204.
- [10] Canisius, F., Fernandes, R., 2012. Evaluation of the information content of Medium Resolution Imaging Spectrometer (MERIS) data for regional leaf area index assessment. *Remote Sens. Environ.* 119, 301–314
- [11] Ottavian, M., Facco, P., Fasolato, L., Barolo, M., 2012. Multispectral data classification using similarity factors. *Chemom. Intell. Lab. Syst.* 118, 13–23.
- [12] Ulrici, A., Serranti, S., Ferrari, C., Cesare, D., Foca, G., Bonifazi, G., 2013. Efficient chemometric strategies for PET–PLA discrimination in recycling plants using hyperspectral imaging. *Chemom. Intell. Lab. Syst.* 122, 31–39.
- [13] Terra, L.A., Poppi, R.J., 2014. Monitoring the polymorphic transformation on the surface of carbamazepine tablets generated by heating using near-infrared chemical imaging and chemometric methodologies. *Chemom. Intell. Lab. Syst.* 130, 91–97.
- [14] Esquerre, C., Gowen, A.A., Burger, J., Downey, G., O'Donnell, C.P., 2012. Suppressing sample morphology effects in near infrared spectral imaging using chemometric data pre-treatments. *Chemom. Intell. Lab. Syst.* 117, 129–137.
- [15] Offroy, M., Roggo, Y., Duponchel, L., 2012. Increasing the spatial resolution of near infrared chemical images. *Chemom. Intell. Lab. Syst.* 117, 183–188.
- [16] Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, J.M., 2007. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, *Trends Food Sci. Tech.* 18, 590–598.
- [17] Alvarez-Jubete, L., Mishra, J., Jones, I., Cullen, P.J., Sullivan, C., 2013. Feasibility of near infrared chemical imaging for pharmaceutical cleaning verification, *J. Near Infrared Spec.* 21, 173–182.
- [18] Vajna, B., Pataki, H., Nagy, Z., Farkas, I., Marosi, G., 2011. Characterization of melt extruded and conventional Isoptin formulations using Raman chemical imaging and chemometrics, *Int. J. Pharm.* 419, 107–113.
- [19] Carneiro, R.L., Poppi, R.J., 2014. Infrared imaging spectroscopy and chemometric tools for in situ analysis of an imiquimod pharmaceutical preparation presented as cream, *Spectrochim Acta A.* 118, 215–220.
- [20] Edelman, G., Lopatka, M., Aalders, M., 2013. Objective color classification of ecstasy tablets by hyperspectral imaging, *J. Forensic. Sci.* 58, 881–866.
- [21] Bishop, C.M., *Pattern recognition and machine learning*, Springer, New York,

2006.

- [22] Toher, D., Downey, G., Murphy, T.B., 2007. A comparison of model-based and regression classification techniques applied to near infrared spectroscopic data in food authentication studies. *Chemom. Intell. Lab. Syst.* 89, 102–115.
- [23] Kaneko, H., Funatsu, K., 2013a. Estimation of predictive accuracy of soft sensor models based on data density. *Chemom. Intell. Lab. Syst.* 128, 111–117.
- [24] Crammer, K., Singer, Y., On the algorithmic implementation of multi-class SVMs, *JMLR*, 2, 265–292.
- [25] Arakawa, M., Yamashita, Y., Funatsu, K., 2011. Genetic algorithm-based wavelength selection method for spectral calibration. *J. Chemom.* 25, 10–19.
- [26] Kim, S., Kano, M., Nakagawa, H., Hasebe, S., 2011. Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection. *Int. J. Pharm.* 421, 269–274.
- [27] Kaneko, H., Funatsu, K., 2012. A new process variable and dynamics selection method based on a genetic algorithm-based wavelength selection method. *AIChE J.* 58, 1829–1840.
- [28] Kaneko, H., Funatsu, K., 2013b. Nonlinear regression method with variable region selection and application to soft sensors. *Chemom. Intell. Lab. Syst.* 121, 26–32.
- [29] Escobar, M.S., Kaneko, H., Funatsu, K., 2014. Flour concentration prediction using GAPLS and GAWLS focused on data sampling issues and applicability domain. *Chemom. Intell. Lab. Syst.* 137, 33–46.
- [30] Kaneko, H., Funatsu, K., 2014. Applicability domain based on ensemble learning in classification and regression analyses. *J. Chem. Inf. Model.* 54, 2469–2482.
- [31] Leardi, R., 2011. Genetic algorithms in chemometrics and chemistry: a review. *J. Chemometr.* 15, 559–569.
- [32] Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58, 109–130.
- [33] Houck, C.R., Joines, J.A., Kay, M.G., 1995. A genetic algorithm for function optimization: A Matlab implementation; NCSU-IE TR 95-09; Meta-heuristic Research and Applications Group: North Carolina State University: Raleigh, NC.
- [34] Savitzky, A., Golay, M.J.E., 1964, Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639.
- [35] Luo, J. Ying, K., Bai, J., 2005. Savitzky–Golay smoothing and differentiation filter for even number data. *Signal Process.* 85, 1429–1434.
- [36] Yoshimura, N., Takayanagi, M., 2012. Chemometrics calculations with Microsoft Excel (5), *J. Comput. Chem., Jpn.* 11, 149–158. (in Japanese)

Figures captions

Figure 1. Chromosome in GAWLS.

Figure 2. Basic concept of the proposed method.

Figure 3. Classification results of a test hyperspectral image for kNN. Each of pixels shown as points is classified using kNN. To show the difference of points clearly, an interval between each pixel is larger than the actual one.

Figure 4. ROC curves for kNN. Blue dashed line, green dotted line and red line represent the results without A, B and C, respectively.

Figure 5. Wavelength selection results for the proposed method.

Figure 6. Frequency of selection made for all the five NW values for the proposed method.

Table 1. Prediction results. Each index is calculated using the test data set, which is used as an external validation set that is separate from the training data set. All values are average in 10 calculations for GAWLS-kNN.

	# of input variables	AR	AUC (A is unknown)	AUC (B is unknown)	AUC (C is unknown)
kNN	253	0.997	0.997	0.997	0.999
GAWLS-kNN NW = 1	25	0.997	0.994	0.999	0.993
GAWLS-kNN NW = 2	50	0.995	0.994	0.994	0.994
GAWLS-kNN NW = 3	53	0.995	0.994	0.992	0.995
GAWLS-kNN NW = 4	58	0.996	0.995	0.992	0.996
GAWLS-kNN NW = 5	64	0.995	0.995	0.994	0.995

417

The first numbers of
the selected wavelengths

The wavelengths whose numbers
are 1265 through 1283 and
2057 through 2069 are selected.

The widths of the selected
wavelengths from the left numbers.

418

419

420 Figure 1. Chromosome in GAWLS.

421

Figure 2. Basic concept of the proposed method.

428

429

430

431 Figure 3. Classification results of a test hyperspectral image for kNN. Each of pixels
432 shown as points is classified using kNN. To show the difference of points clearly, an
433 interval between each pixel is larger than the actual one.

434

435

436

437

438 Figure 4. ROC curves for kNN. Blue dashed line, green dotted line and red line
439 represent the results without A, B and C, respectively.

440

441

(a) NW = 1

(b) NW = 2

(c) NW = 3

Figure 5. Wavelength selection results for the proposed method.

Figure 6. Frequency of selection made for all the five NW values for the proposed method.