Classification of Food Powders using Handheld NIR Spectrometer

Hanjong You
Dept. of Secured Smart Electric Vehicle
Kookmin University
Seoul, Korea 02707
open803@kookmin.ac.kr

Youngsik Kim
Stratio, Inc.
San Jose, CA 95131
youngsik.kim@stratiotechnology.com

Jae Hyung Lee
Stratio, Inc.
San Jose, CA 95131
jaehyung.lee@stratiotechnology.com

Sunwoong Choi
Dept. of Secured Smart Electric Vehicle
Kookmin University
Seoul, Korea 02707
schoi@kookmin.ac.kr

Abstract**—** Near Infrared (NIR) spectroscopy is attracting much attention as a fast and non-destructive method for analyzing materials. However, general laboratory NIR spectrometers are not suitable for use at home or in the field due to their sizes and prices. In this paper, we classify 8 common food powders using handheld NIR device with range of 450-1000nm. We develop a machine learning model using spectral data of 8 food powders by 3 supervised classification methods. We have obtained 100% correct classification results and it shows the possibility of ingredient analysis of various foods using handheld NIR device.

Keywords—near infrared spectroscopy; handheld divces; food powder; supervised classification method

I. Introduction

NIR spectroscopy is powerful tool for analyzing materials. It can be used for identification of ingredient of food. NIR spectroscopy first was applied in agricultural applications by Norris to measure the moisture in grain [1]. It is a fast and non-destructive method compared to existing destructive material analysis methods. After that, it used for the analysis of various products and foods [2]. In addition, various chemometrics and pretreatment techniques have been developed and many experiments proceed such as identification of genetically modified organisms (GMO) foods [3][4].

Industrial or laboratory NIR spectrometers are excellent in performance. However, it is not suitable for use at home or in the field, due to their size and price. Therefore handheld NIR devices are being actively developed and verified[5]. The handheld NIR devices are inexpensive compared with laboratory NIR spectrometer and are available to use at home or in the field.

In this paper, we classify food powders using handheld NIR device with 3 supervised classification methods. The result of classification analyzes to verify validation. It shows possibility about analyzing ingredient of foods at home.

The rest of the paper is as follows: First we introduce a NIR spectroscopy for identification of ingredient of food, and the equipment, food powders and supervised classification methods used in experiment. And we describe our training and validation process in Section III. We analyze the results of 3 supervised classification methods in Section IV. Finally, we present the possibility of classification food powders using handheld NIR device in Section V.

II. Materials and Methods

A. Handheld NIR Device

We use handheld NIR device from Stratio, Inc. (www.stratiotechnology.com) called LinkSquare^{TM}. The LinkSquare^{TM} is a Silicon (Si)-based VIS/NIR spectrometer. It is a lot affordable than the NIR spectrometers we used to see from the laboratory. And it has two light sources and measures the range of 450-1000nm [6]. Stratio also provides the Software Development Kit (SDK) for users to collect their own data and to develop their own apps. The actual image of the device and a graph obtained using the SDK are shown in Fig. 1.

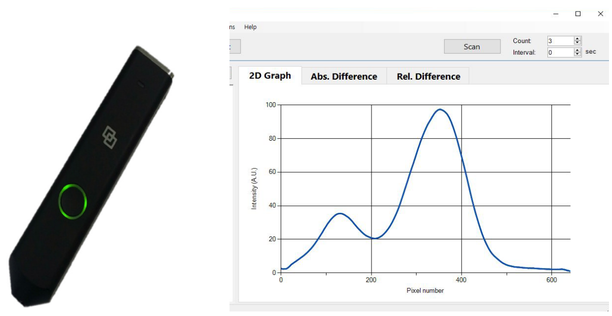
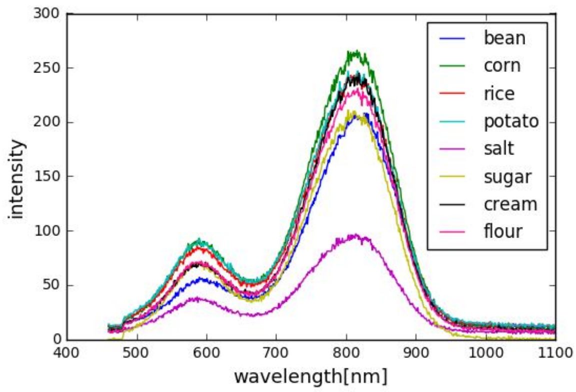


Fig. 1. LinkSquare^{TM} and SDK

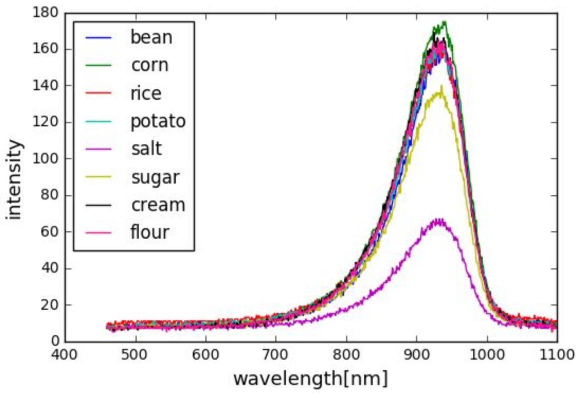
B. Food Powders

In this paper, we consider 8 food powders which are commonly used and visually indistinguishable: sugar, salt, cream, flour, corn, rice, bean and potato powder.

We measure each 8 food powders 150 times using LinkSquare^{TM}. The measured spectral data to each light source are shown in Fig. 2.



(a) Spectral data using LED



(b) Spectral data using BULB

Fig. 2. NIR spectra of food powders

III. Classification of Food powders

A. *Supervised Classification Methods*

We use the 3 supervised classification methods are in general use in this experiment [7].

1) SVM (Support Vector Machine): SVM is a useful machine learning technique for data classification. The advantages of SVM are effective in high dimensional spacese and efficient in cases where number of dimensions is greater than the number of samples [8].

2) KNN(k Nearest Neighbors): The k nearest neighbor method is one of the simplest machine learning algorithms. KNN stores all training data and calculates the distance between the neighbors and test data. Then KNN finds the k closest training data, and it predicts many number of labels as test label among the k neighbors [9].

3) RF(Random Forest): Breiman proposed random forest which consists of serveral decision tree and called a kind of ebsemble learning[10]. And radom forest has a number of trees as an important hyperparameter. If the number of trees is small, training speed is fast but the accuracy is low. The larger the number of trees, the higher the accuracy but training speed is slow.

The 3 supervised classification methods verify and calculate accuracy of each learning model. And we use the result to find the most appropriate supervised classification method for classification of food powders.

B. *Training & Validation Method*

Fig. 3 shows training and validation process. We divide 1200 scan data into 1000 training sets and 200 validation sets. Then we develop learning models with training set using 3 supervised classification methods. We obtain accuracy of each models with validation sets.

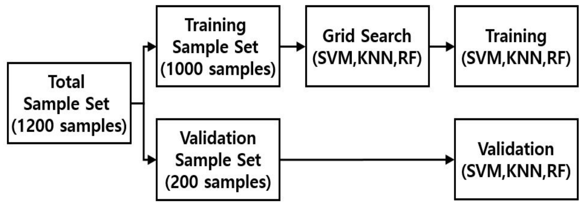


Fig. 3. Training and Validation Process

C. *Optimal Parameter Selection*

We use the grid-search function of scikit-learn to optimize parameters of 3 supervised classification methods before training [8]. The grid-search function is a method to find for optimization parameters of machine learning. First, we set the parameter range and use the training set to find the appropriate parameters.

We set the parameter ranges of SVM, KNN, and RF as in Table I. Table I shows optimal parameters that found for each supervised classification method

TABLE I. Parametes Opimization

Classification	Parameter	Range	Selected Optimal Parameter
SVM	Kernel	Linear, RBF	RBF
	C	1, 10, 100, 1000	10
	Gamma	0.001, 0.0001	0.0001
KNN	Algorithm	auto, ball_tree, kd_tree, brute	auto
	n_neighbors	1 ~ 100	1
RF	n_estimators	1 ~ 200	23

IV. CLASSIFICATION RESULTS

We verify models by learning 3 supervised classification methods using validation set. The validation results are shown in Table II. And the Receiver Operating Characteristic (ROC) curves are represented in Fig. 4.

TABLE II. CLASSIFICATION RESULTS

Food Powders	Accuracy		
	SVM	**KNN**	**RF**
salt	100%	100%	100%
sugar	100%	100%	100%
cream	100%	100%	80%
flour	100%	90%	35%
bean	100%	100%	100%
corn	100%	100%	100%
rice	100%	100%	100%
potato	100%	95%	90%
Average	**100%**	**98.12%**	**87.5%**

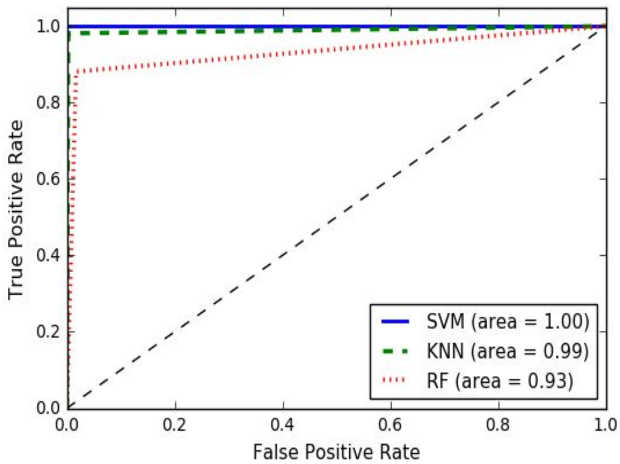


Fig. 4. ROC curve

In experiment, we obtain 100% correct classification using the SVM model. And KNN and RF model show 98% and 87% of accuracy. The ROC curves show Area under the Curve (AUC) larger than 0.94. The ROC curve is used as one of the machine learning evaluation tools [11]. In general, if the AUC is more than 0.90, it is sufficiently good classification examination [12].

In summary, all supervised classifications have showed high accuracy of higher than 85%. It confirms the possibility of the classification of food powders using handheld NIR device.

V. CONCLUSION

We present the possibility of ingredient analysis of various foods using handheld NIR device. And we can use this powerful tool to identify various things that we are difficult to identify such as GMO food and country of origin instead of existing laboratory NIR spectrometers. As the handheld NIR devices develop further, they can be used for more various purposes.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2016-0-00080, Germanium on Silicon based Sensor which covers visible spectrum to Short Wavelength Infra-Red range (400-1600nm) for Materials Identification and Application System Development)

REFERENCES

[1] Norris and Karl H. "Design and development of a new moisture meter," Agric. Eng 45.7, pp. 370-372, 1964.

[2] Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I. and Lammertyn, J. "Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review," Postharvest biology and technology 46.2, pp. 99-118, 2007.

[3] Xie, Lijuan, Yibin Ying and Tiejin Ying. "Combination and comparison of chemometrics methods for identification of transgenic tomatoes using visible and near-infrared diffuse transmittance technique," Journal of Food Engineering 82.3, pp. 395-401, 2007.

[4] Dong-Jin Kang, Ji-Young Moon, Dong-Gil Lee and Seong-Hun Lee. “Identification of the geographical origin of cheonggukjang by using fourier transform near-infrared spectroscopy and energy dispersive X-ray fluorescence spectrometry,” Korean journal of food science and technology 48.5, pp. 418-423, 2016.

[5] Das, A. J., Wahi, A., Kothari, I. and Raskar, R. "Ultra-portable, wireless smartphone spectrometer for rapid, non-destructive testing of fruit ripeness," Scientific Reports 6, 2016.

[6] http://www.linksquare.io.

[7] Li, Juan and Chandima Fernando. "Smartphone-based personalized blood glucose prediction," ICT Express 2.4, pp.150-154, 2016.

[8] http://scikit-learn.org.

[9] Fukunaga, Keinosuke and Patrenahalli M. Narendra. "A branch and bound algorithm for computing k-nearest neighbors," IEEE transactions on computers 100.7, pp. 750-753, 1975.

[10] Liaw, Andy and Matthew Wiener. "Classification and regression brandomForest," R news 2.3, pp. 18-22, 2002.

[11] Bradley and Andrew P. "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern recognition 30.7, pp. 1145-1159, 1997.

[12] Hanley, James A. and Barbara J. McNeil. "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology 143.1, pp. 29-36, 1982.