

Food Powder Classification Using a Portable Visible-Near-Infrared Spectrometer

Hanjong You¹ · Youngsik Kim² · Jae-Hyung Lee² · Byung-Jun Jang³ · Sunwoong Choi^{1,3,*}

Abstract

Visible-near-infrared (VIS-NIR) spectroscopy is a fast and non-destructive method for analyzing materials. However, most commercial VIS-NIR spectrometers are inappropriate for use in various locations such as in homes or offices because of their size and cost. In this paper, we classified eight food powders using a portable VIS-NIR spectrometer with a wavelength range of 450–1,000 nm. We developed three machine learning models using the spectral data for the eight food powders. The proposed three machine learning models (random forest, k-nearest neighbors, and support vector machine) achieved an accuracy of 87%, 98%, and 100%, respectively. Our experimental results showed that the support vector machine model is the most suitable for classifying non-linear spectral data. We demonstrated the potential of material analysis using a portable VIS-NIR spectrometer.

Key Words: Classification, Food Powder, Machine Learning, Near Infrared Spectroscopy, Portable VIS-NIR Spectrometer.

I. INTRODUCTION

A spectrometer is a sensing device for classifying various materials based on the interactions between electromagnetic waves and the material. Radiated electromagnetic waves can be absorbed or reflected by a material. Therefore, the spectrum of the reflected electromagnetic wave as a function wavelength can be considered a fingerprint of the material. Among various spectrometers, a visible-near-infrared (VIS-NIR) spectrometer is useful for analyzing materials and can be used to identify the constituents of food. VIS-NIR spectroscopy was first applied in agriculture by Norris to measure the moisture in grain [1]. Various spectrometers and pretreatment techniques have been developed for analyzing the constituents of various materials and foods [2–4]. Re-

cently, food products containing genetically modified organisms (GMO) have been studied using NIR spectroscopy [5].

Industrial or laboratory VIS-NIR spectrometers have excellent performance. However, these spectrometers are not suitable for use in various locations such as in homes or offices because of their size and cost. Therefore, portable VIS-NIR spectrometers are being actively developed and validated [6].

In this paper, we classify eight food powders using a portable VIS-NIR spectrometer with three supervised classification methods that are generally used. Our experimental results demonstrate the potential for analyzing food ingredients using a portable VIS-NIR spectrometer.

The rest of the paper is organized as follows. In Section II, we introduce the VIS-NIR spectroscopy for the identifi-

Manuscript received September 1, 2017 ; Revised October 16, 2017 ; Accepted October 17, 2017. (ID No. 20170901-044J)

¹Department of Secured Smart Electric Vehicle, Kookmin University, Seoul, Korea.

²Stratio Inc., San Jose, CA, USA.

³Department of Electrical Engineering, Kookmin University, Seoul, Korea.

*Corresponding Author: Sunwoong Choi (e-mail: schoi@kookmin.ac.kr)

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© Copyright The Korean Institute of Electromagnetic Engineering and Science. All Rights Reserved.

cation of food constituents and discuss the disadvantages of existing laboratory VIS-NIR spectrometers. The portable VIS-NIR spectrometer, food powders, and supervised classification algorithms used in the experiment are then explained. In Section III, we describe our machine learning process. We analyze the results of the three machine learning algorithms used and the effect of the training set size in Section IV. Finally, we conclude with a discussion of our results.

II. MATERIALS AND METHODS

1. Portable VIS-NIR Spectrometer

We use a portable VIS-NIR spectrometer from Stratio Inc. (www.stratiotechnology.com) called LinkSquare. LinkSquare in Fig. 1 is a Silicon (Si)-based VIS-NIR spectrometer that is significantly more affordable than the NIR spectrometers typically found in the laboratory. This spectrometer has two light sources, white LED and BULB, and measures within the wavelength range of 450–1,000 nm [7]. Table 1 provides the detailed specifications of LinkSquare.

2. Food Powders and VIS-NIR Spectra

In this paper, we evaluate eight common food powders that are visually indistinguishable: salt, sugar, cream, flour, bean, corn, rice, and potato powder. Fig. 2 shows the eight food powders selected.

We measure the eight food powders using the portable VIS-NIR spectrometer. The process of spectral data acquisition



Fig. 1. LinkSquare and the process of spectral data acquisition.

Table 1. Specifications of LinkSquare

Company	Stratio Inc.
Name	LinkSquare
Measure wavelength range	450–1,000 nm
Size	114 mm × 23.9 mm × 23.9 mm (4.5 in × 9 in × 9 in)
Weight	57 g/2 oz
Battery (active)	Approximately 1,000 scans
Battery (idle)	> 24 hr



Fig. 2. Eight food powders.

tion as illustrated in Fig. 1, is conducted in a constant condition of ambient illumination and measuring angle. The spectral data obtained with each light source of the spectrometer are shown in Fig. 3.

III. CLASSIFICATION OF FOOD POWDERS

1. Supervised Classification Methods

In this paper, we use three supervised classification methods for machine learning: support vector machine (SVM), k -

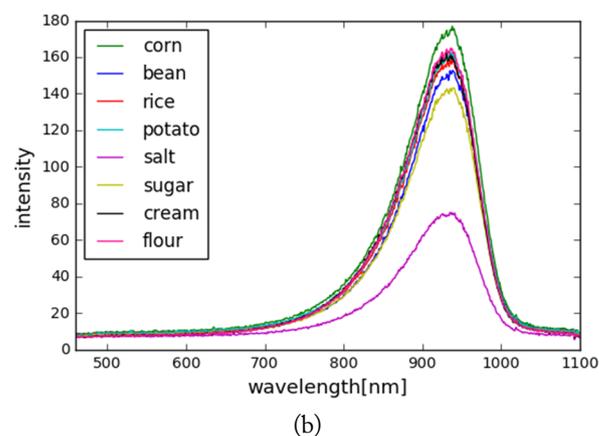
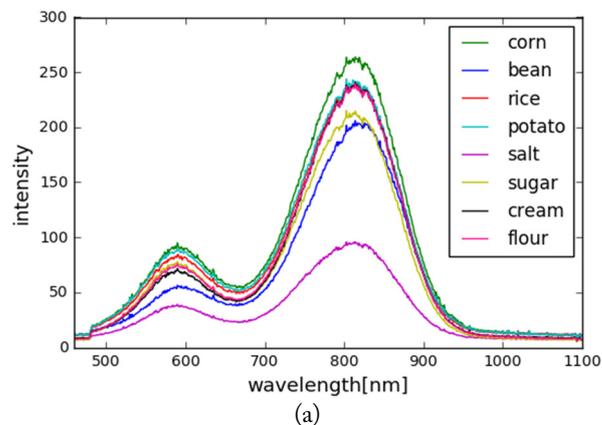


Fig. 3. Spectral data: (a) white LED, (b) BULB.

nearest neighbors (kNN), and random forest (RF). SVM is one of the standard methods of classification and is most effective when the number of dimensions is the greater than the number of samples [8]. kNN utilizes a simple algorithm based on the distance between the training data and test data. It defines the nearby k training data as a neighbor and predicts the greatest number of labels among the k neighbors as the test label [9]. Proposed by Breiman, RF consists of several decision trees, and it is also described as ensemble learning [10]. RF uses the number of trees as an important hyperparameter. If the number of decision trees is small, the training speed will be fast but the accuracy will be low. Conversely, the larger the number of decision trees is, the higher the accuracy but the slower the training speed.

All machine learning methods are implemented in Python with the use of numpy, scipy, and scikit-learn [8].

2. Training and Validation Method

Fig. 4 shows the machine learning process in this experiment. We divide the total sample set (960 samples) into a training sample set (800 samples) and a validation sample set (160 samples). We then design the machine learning models with the training sample set using the three supervised classification methods. We demonstrate the performance of each model with the validation sample set.

3. Optimal Parameter Selection

We use the grid-search function to find the optimal parameters for the three machine learning methods. We set the suitable parameter ranges for SVM, kNN, and RF and then, find the optimal parameters for machine learning through repetitive experiments using the grid-search function. The detailed parameters for the experiment are given in Table 2.

SVM can use a kernel function, and thus we consider the linear and radial basis function (RBF) kernel. The RBF kernel has a parameter gamma, which defines how much influence a single training example has. Parameter C is called the penalty parameter, which controls the tradeoff between margin maximization and error minimization. The performance of kNN varies on the basis of parameter k , called n_neighbors in scikit-learn. In general, a larger k suppresses the effects of noise, but makes the classification boundaries

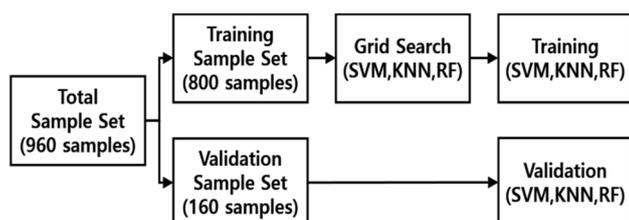


Fig. 4. Training and validation process.

Table 2. Parameter optimization

Classification method	Range	Selected optimal parameter
SVM		
Kernel	Linear, RBF	RBF
Gamma	0.001, 0.0001	0.0001
C	1, 10, 100, 1000	10
kNN		
n_neighbors	1–100	1
RF		
n_estimators	1–200	23

smoother. RF models with different parameter n_estimators which are the number of trees in the forest are evaluated.

IV. CLASSIFICATION RESULTS

1. Results and Confusion Matrix

We verify the three machine learning models using the validation sample set. We evaluate the performance of the classification in terms of accuracy, recall, precision and F_1 score [11].

Table 3 shows the results of the experiment. We observe that the three machine learning methods almost successfully classify all eight food powders. RF, kNN, and SVM achieve an accuracy of 87%, 98%, and 100%, respectively. SVM shows high performance because it transforms the non-linear spectral data into the maximum-margin hyperplane.

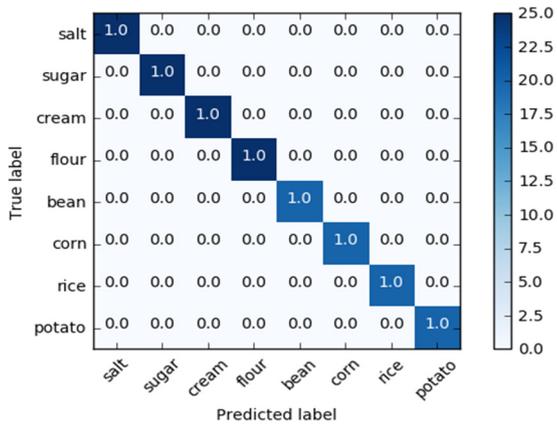
We further investigate the confusion matrices for the three machine learning methods as shown in Fig. 5. Although kNN and RF have high accuracy, they fail to accurately classify some food powders. In particular, 72% of the flour powder is misidentified as rice powder.

2. Size of the Training Sample Set

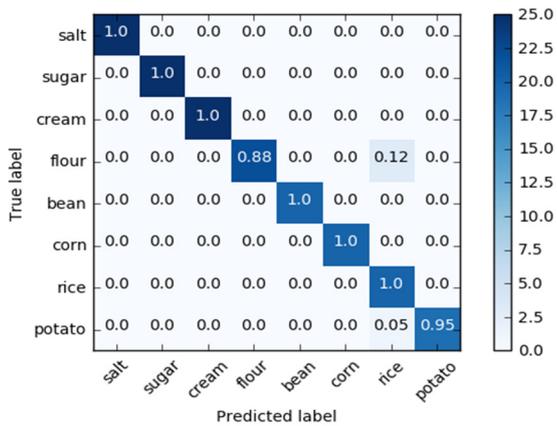
Next, we investigate the effect of the training sample set size for efficient machine learning training. We test iteratively by changing the size of the training sample set. Fig. 6 shows the effect of the training sample set size on the classification performance. The classification accuracy enhances as the training sample set size increases and plateaus after a certain size. We obtain as many training samples as we obtain to guarantee the classification performance of SVM, but

Table 3. Classification results

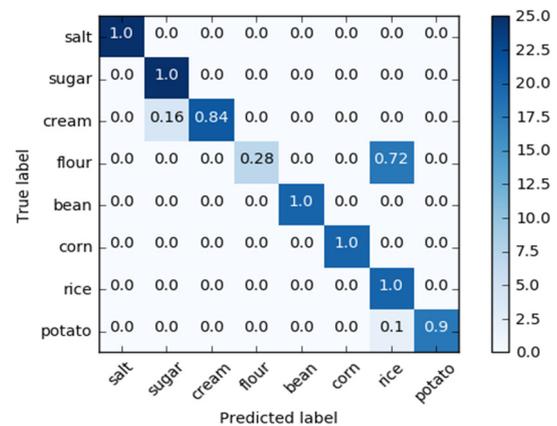
	SVM (%)	kNN (%)	RF (%)
Accuracy	100	97.7	86.5
Recall	100	97.7	86.5
Precision	100	97.7	86.4
F_1 score	100	97.7	86.5



(a)



(b)



(c)

Fig. 5. Confusion matrix: (a) SVM, (b) kNN, and (c) RF.

RF requires more training samples than what we have obtained.

V. CONCLUSION

We present the possibility of converging VIS-NIR spectroscopy and machine learning in this paper. Eight food powders are classified using a portable VIS-NIR spectrometer with three supervised classification methods. The

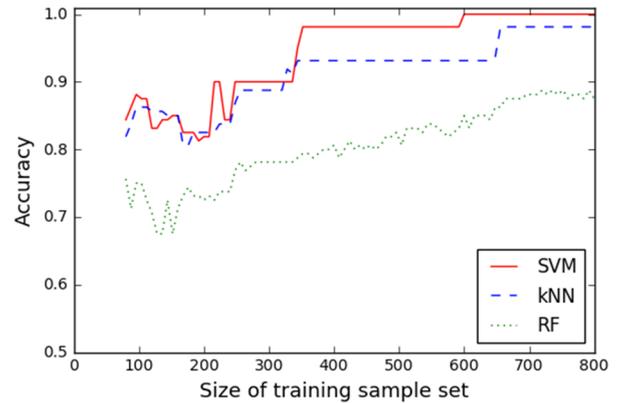


Fig. 6. Effect of the training sample set size.

successful classification results for the eight food powders show the feasibility of using a portable VIS-NIR spectrometer for analyzing food ingredients. As portable VIS-NIR devices develop further, they can be used for more varied purposes.

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2016-0-00080, Germanium on silicon based sensor which covers visible spectrum to short wavelength infra-red range (400–1600 nm) for materials identification and application system development).

REFERENCES

- [1] K. H. Norris, "Design and development of a new moisture meter," *Agricultural Engineering*, vol. 45, no. 7, pp. 370–372, 1964.
- [2] B. G. Osborne and T. Fearn, "Near-infrared spectroscopy in food analysis," BRI Australia Ltd, North Ryde, Australia, 2004.
- [3] D. J. Kang, J. Y. Moon, D. G. Lee, and S. H. 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*, vol. 48, no. 5, pp. 418–423, 2016.
- [4] T. D. Kim, S. H. Lee, K. J. Baik, B. J. Jang, and K. H. Jung, "Classification of tablets using a handheld NIR/visible-light spectrometer," *The Journal of Korean Institute of Electromagnetic Engineering and Science*, vol. 28, no. 8, pp. 628–635, 2017.
- [5] L. Xie, Y. Ying, and T. Ying, "Combination and comparison of chemometrics methods for identification of

transgenic tomatoes using visible and near-infrared diffuse transmittance technique," *Journal of Food Engineering*, vol. 82, no. 3, pp. 395–401, 2007.

- [6] A. J. Das, A. Wahi, I. Kothari, and R. Raskar, "Ultra-portable, wireless smartphone spectrometer for rapid, non-destructive testing of fruit ripeness," *Scientific Reports*, vol. 6, article no. 32504, 2016.
- [7] LinkSquare, <http://www.linksquare.io>.
- [8] scikit-learn machine learning in Python, <http://scikit-learn.org>.

- [9] F. Keinosuke and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Transactions on Computers*, vol. 100, no. 7, pp. 750–753, 1975.
- [10] L. Andy and M. Wiener, "Classification and regression brandomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [11] P. D. Martin, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

Hanjong You



received his B.S. degree in Electrical Engineering from Kookmin University, Seoul, Korea, in 2017. He is currently working toward his M.S. degree at the Department of Secured Smart Electric Vehicle, Kookmin University. His research interests include machine learning and embedded system.

Byung-Jun Jang



received his B.S., M.S., and Ph.D. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1990, 1992, and 1997, respectively. From 1995 to 1999, he worked for LG Electronics in Seoul, where he developed code-division multiple access and digitally enhanced cordless telecommunication RF modules. From 1999 to 2005, he worked at the Electronics and Telecommunications Research Institute, Daejeon, Korea, where he performed research in the fields of satellite RF components and monolithic microwave integrated circuits. In 2005, he joined Kookmin University, Seoul, where he is currently with the Department of Electrical Engineering. His current research interests are RF circuit design, radio frequency identification system design, wireless power transfer system design, frequency interference modeling and spectrum engineering, and wireless sensor design.

Youngsik Kim



received his B.S. degree in electrical engineering in 2006 from Seoul National University and his M.S. and Ph.D. degrees in electrical engineering in 2009 and 2017, respectively, from Stanford University, Stanford, CA. He is a co-founder, vice president, and systems engineer of Stratio. He is an expert in digital circuit design, software development, and integration of sensors and system.

Jae-Hyung Lee



received his B.S. degree in electrical engineering in 2004 from Seoul National University and his M.S. and Ph.D. degrees in electrical engineering in 2008 and 2014, respectively, from Stanford University, Stanford, CA. His research interests include photon-enhanced thermionic energy converters, Germanium-based infrared image sensors, SiC-based power devices, low-cost handheld spectrometers, and various wafer bonding techniques. He is currently the CEO and co-founder of Stratio (www.stratiotechnology.com), where he is working on a smart handheld spectrometer, LinkSquare (linksquare.io), and low-power high-resolution Germanium-based infrared image sensors.

Sunwoong Choi



received his B.S., M.S., and Ph.D. degrees in electrical and computer engineering from Seoul National University, Korea, in 1998, 2000, and 2005, respectively. He worked for Samsung Electronics from 2005 to 2007. Since 2007, he has been a faculty member in the School of Electrical Engineering, Kookmin University, Korea. His research interests include MAC and routing protocols for wireless networks, network resource management, and machine learning.