

# Visible/near-infrared Spectroscopy as a Novel Technology for Nondestructive Detection of *Escherichia coli* ATCC 8739 in Lettuce Samples

Sahar Rahi<sup>\*1</sup>, Hossein Mobli<sup>1</sup>, Bahareh Jamshidi<sup>\*2</sup>, Aslan Azizi<sup>2</sup>, Mohammad Sharifi<sup>1</sup>

<sup>1</sup>Department of Agricultural Machinery Engineering, University of Tehran, Tehran, Iran

<sup>2</sup>Agricultural Engineering Research Institute, Agricultural Research Education and Extension Organization (AREEO), Karaj, Iran

\* Corresponding author

## Abstract

Microbial contamination of vegetables as a major safety challenge is known in recent years. Detection of *Escherichia coli* (*E. coli* ATCC 8739) in vegetables especially lettuce is very important issue for post-harvest management in Iran. Non-destructive detection and evaluation of *E. coli* contamination in fresh lettuce leaf as a main aim of this work was explored using Visible Near-infrared spectroscopy (Vis/NIR). Reflectance spectra in intransmittance mode were measured using a spectrometer at the range of 350–1100 nm, and 200 fresh lettuce leaves given three different treatments of *E. coli* solution (0.1, 0.2 and 0.3 ml) were used for spectra measurements and total *E. coli* determination. Classification of lettuce samples based on different concentration of *E. coli* solution was performed by Partial least Squares Discriminant Analysis (PLS-DA) with different pre-processing methods into the two different groups of “safe” and “unsafe” samples. The best model was recommended using standard normal variation (SNV) + second derivative (D2) pre-processing methods with the minimum standard deviation of cross-validation (SECV = 0.256). Besides, a good correlation ( $R^2$  of 0.93) between Vis/NIR spectral data and the presence of *E. coli* contamination proved the possibility of Vis/NIR spectroscopy for microbial detection and evaluation in lettuce.

**Keywords:** Near-infrared (NIR) spectroscopy; Lettuce; Microbial detection; Food safety

## 1. Introduction

*Escherichia coli* (abbreviated as *E. coli*) is a foodborne zoonotic pathogen that found in the environment, foods, and intestines of people and animals. One of the huge group of microorganisms is *E. coli*. Most types of *E. coli* are harmless. Nonetheless, some kinds of *E. coli* can cause serious illnesses like diarrhea, urinary infections, pneumonia, and other illnesses (Jamshidi et al., 2016).

There have been a number of outbreaks reported universally that have been related with the percentage of consumption of vegetables in leafy green contaminated by *E. coli* (Jamshidi et al, 2015; Nicolaï et al., 2014; Zhang et al, 2010). Besides, lettuce is probably presented as the rank one of consumption in leafy green vegetables, which cultivated all over the world (Lu et al., 2013). In Iran, all lettuce types have about 528.7K tons (yearly) of the total volume of leafy vegetables produce in the world (Lu et al., 2013). Other than this, because lettuces like other fresh vegetables are served raw, it is possibly to be spoilage by *E. coli*.

Most of the techniques for microbial evaluation are destructive, time-consuming, require skilled laborers, lab equipment or materials which make it expensive (Jamshidi et al., 2015). Thus, it is important to develop non-destructive, accurate, quick and most important valid detection method for microbial contamination.

In recent years, there have been growing research works in developing non-destructive methods especially optical-based methods for evaluation external and/or internal quality features of different food products.

Spectroscopy and spectral imaging technologies have been evaluated for non-destructive quality evaluation of food products. These methods play important responsibilities in post-harvest management. In addition, these up-to-date techniques are also increasingly being explored and recommended for microbial detection and evaluation of food quality in both physical and chemical features (Foley et al., 1998).

All of the different types of spectroscopy techniques can be accomplished in three reflectance, transmission, or fluorescence modes. One of the most common mode is reflectance and usually use in the Vis/NIR (400-1000 nm) or NIR (1000-1700 nm) range. This range of wavelength can be utilized to detect invisible defects, contaminants like microbial, biological, fungal infection, and physical quality features of wide ranges of products like fruits, vegetables and meat products (André & Lawler, 2003; Williams & Norris, 2001). Fluorescence spectroscopy is utilized in the dairy production (Gowen et al, 2007) to detect facial contamination in food, or botanical tissue which contains chlorophyll (Wu & Sun, 2013; Panagou et al., 2014) . Online assessment of principal concentrations and internal defects of foods also applied by transmission imaging (Park et al., 2011). A main disadvantage of spectroscopy is that, due to undesired signals or noises, the qualitative/quantitative analysis of the original spectral data requires different pre-processing methods (Wang et al., 2014), to decrease numerous useless signals (Zoubir, 2012).

Spectroscopy is used to evaluate the microbial properties of different products like beef, Pork, chicken, fish, Milk, cheese, and apple juice (Stewart et al, 2012). Reviews have shown that fewer studies have been done on vegetables.

In the investigation reported using Fourier Transform Infrared (FTIR) spectroscopy for evaluation of being *Escherichia coli* K-12 in baby spinach. The results showed that FTIR could detect and evaluate *E. coli* K-12 in extracts of baby spinach with approximate detection of 100 CFU/mL in five minute (Wang et al, 2010). Using near infrared (NIR) spectroscopy to quantify the amount of bacterial contamination in fresh cabbage investigated by Sutiluk et al (2008) Two calibration equations by PLS method were developed: both from spectra of the saline solutions which were used to wash the entire cabbage samples and only the outer surface of the sample, respectively. The two equations using the wavelength region were equal. Also, accurate SEPs = 0.46 and CFU = 0.44 log (g<sup>-1</sup>) were reported (Sutiluk et al, 2008).

Based on our knowledge, there is no announced investigation on detection of *E. coli* infected lettuce by the Vis/NIR spectroscopy technique. The main aims of this investigation were (1) develop a non-destructive and valid method for *E. coli* microbial load detection in lettuce, and (2) safety assessment of lettuce samples based on *E. coli* load using the Vis/NIR spectra.

## 2. Material and methods

### 2.1. Sample preparation (Reference measurements)

Lettuce samples were freshly purchased on day of test from local market in Karaj, Iran. The number of samples were at least 200 lettuce leaves to formulate reliable and suitable models for detecting contamination of the product based on the spectral information of the samples.

According to the main goal of the research, several sample categories were needed: Safe samples (no contamination), and unsafe samples with different concentrations of contamination. The accepted level for *E. coli* in ready to eat lettuce is zero (Iranian National Standards Organization, 2018). *E. coli* ATCC 8739 was obtained from the Research Center for Microbial Technologies and Products, University of Tehran, Tehran, Iran. *E. coli* ATCC 8739 was grown for 24 h in nutrient agar at 37 °C. To acquire 9 levels of *E. coli* solutions which start from 10<sup>1</sup> to 10<sup>9</sup> CFU/mL, Growth medium was consecutively diluted in sterilized 0.1% peptone water. Due to the inappropriate nature of all the microorganisms in the lettuces, the index of total microbial count is mainly considered as a microbial quality index (microbial load), which includes the total number of microbes. The *E. coli* solution injected with the different microbial population in samples with various concentrations (0/1, 0/2, 0/3 ml). The population of *E. coli* after counting was expressed as logarithm of colony forming units (CFU) per milliliter (C0.1 = 4.5 log CFU/mL, C0.2 = 6.67 log CFU/mL and C0.3 = 8.91 log CFU/mL). For each contaminated sample, spectra were taken as a safe sample from another leaf of same lettuce which treated with solution without any microbial population. Microbial tests were carried out in the laboratory of microbiology located at Agricultural Engineering Research Institute, Karaj, Iran.

## 2.2. Vis/NIR measurements

Spectra of the Vis/NIR from samples were acquired at a steady level room temperature and humidity. A green-wave spectrometer (StellarNet Inc, Florida, USA) with a highly sensitive CCD detector with 2048 element (signal to noise ratio: 400:1) was used with fiber optic coupled instruments for measurements in 350-1100 nm wavelength range with 1.6 nm resolution. The data were transferred to the lab top by the SpectraWiz software (StellarNet Inc, Florida, USA).

The spectrometer was calibrated every 20 min. Before spectra acquisition of the both blank and the contaminated lettuce. To calculate the relative of reflectance, spectra of white and dark reference were taken. Spectra were recorded in the interactance mode at the range of 350–1100 nm. In addition, three reflectance spectra were captured at three different areas for each leaf by random, and the averaged spectrum of three spectra as the reflectance ( $R$ ) was used for analysis. Finally, all of the data transformed to absorbance ( $\log 1/R$ ).

## 2.3. Data analysis

Different Pre-processing techniques on Vis/NIR spectra were considered to remove all unnecessary data in the spectra and develop the following classification model or multivariate regression analysis (Elmasry & Nakauchi, 2016). To this end, standard normal variate (SNV) with multiplicative scatter correction (MSC) were used to remove multiplicative and additive scatter effects, respectively. Also, 1st and 2nd derivatives of the spectrum ( $D_1$ ,  $D_2$ ) based on Savitzky–Golay smoothing filtering with 5 points and two polynomial order were accomplished to enhance the spectral resolution (Elmasry et al., 2012). Before any analysis, two segments of the spectrum from the whole wavenumber range were removed to diminish the low value of these signals: 350-450 and 994-1100. To developing the classification models, Principal component analysis method (PCA) was conducted to recognize outliers in data and patterns or relationships between variables and groups.

Partial least squares discriminant analysis (PLS-DA) as a supervised method was used to classify of the lettuce samples into the safe and unsafe groups. In order to avoid bias in selecting subset and estimate the performance of a predictive model, the calibration and validation set was comprised of 75% and 25% of the total samples, respectively. Therefore, the calibration and validation sets were created with 144 (72 safe sample and 72 unsafe samples) and 47 (23 safe samples and 24 unsafe samples) separate samples. The samples belonging to validation set were randomly chosen by picking five rows sample after counting every five sample, making sure that both sets of data at least some sample of each group names was included.

After the PCA and PLS-DA models, full cross validation, sometimes called rotation estimation, or out-of-sample testing, as validation technique was used. Statistically significant of developed PLS-DA classification models was evaluated with standard error of cross-validation (SECV) and the percent correctly classified of validation and calibration sets. In general, an acceptable model should have lowest Root Mean Square Error of Prediction (RMSEP) and Root Mean Squares Error Calibration (RMSEC) with highest coefficient of determination ( $R^2$ ) value. Therefore, the best model was selected taking into account the

standard error of cross-validation (SECV) led to minimum and the percentage variation explained by the regression model ( $R^2$ ) led to maximum.

The Unscrambler software X10.3 (CAMO Software, Oslo, Norway) was used for all analyses.

### **3. Result and Discussion**

#### **3.1. PCA model**

Before any pre-processing methods, Outliers were removed from all of the data to reduce the contribution of potential outliers, which have negative influence on model development [36, 37]. After running principal component analysis (PCA) with the algorithm of nonlinear iterative partial least squares (NIPALS), 9 outliers were detected and removed. To interpret of grouping between samples (patterns and outliers), the score plot, which shows the projected values of the origin data onto the axes of principal component was used. The score plot of first principal component (PC1) and second principal component (PC2) from all lettuce samples is shown in Figure 1. In this investigation, the score plot labels show that PC1 explains 89% and PC2 6% of the total variance in the data.

On one hand, Figure 1 Shows the lettuce samples can be separated into two distinct groups of safe and unsafe lettuce samples with PC1 and PC2, placing unsafe samples in the left side of PC1 and PC2, and safe samples in the right side of PC1 and PC2. However, it is obvious there are some overlap between the samples belonging to safe group and unsafe with 0.1 concentration of microbial solution due to their similarity spectrum ranges. In regards to loading plot (Figure 2), the main wavelength bands were those from 430 to 680 nm. This means that the optimal sensitivity for *E. coli* determination is given in this region. Photosynthetic organisms contain light-absorbing molecules called pigments that absorb only specific wavelengths of visible light could be seen between ranges of 520 to 550 and related to chlorophyll a. (Jamshidi et al., 2015).

On the other hand, the Q-residual-Hotelling's T2 plot indicates the outliers in Figure 3. The Q statistic indicates large leverage and high residual variance of outliers. So, it describes the distance of sample to model. The Hotelling's T2 on the lateral axis describes how each samples is defined by the model. The red line on the Q-residual-Hotelling's T2 plot shows limitation on the 5% level of significance. The samples with a circle around them beyond those limited lines are recognized as the detected outliers. Also, the plot was conducted with the third PC where total residual variance goes to zero for with minimum of components which could be used in PCA model (Aske et al., 2001).

Figure 1: The score plot Score plot of first principal component (PC1) and second principal component (PC2) from all lettuce samples as results of principal component analysis (The samples with a circle around them are outliers).

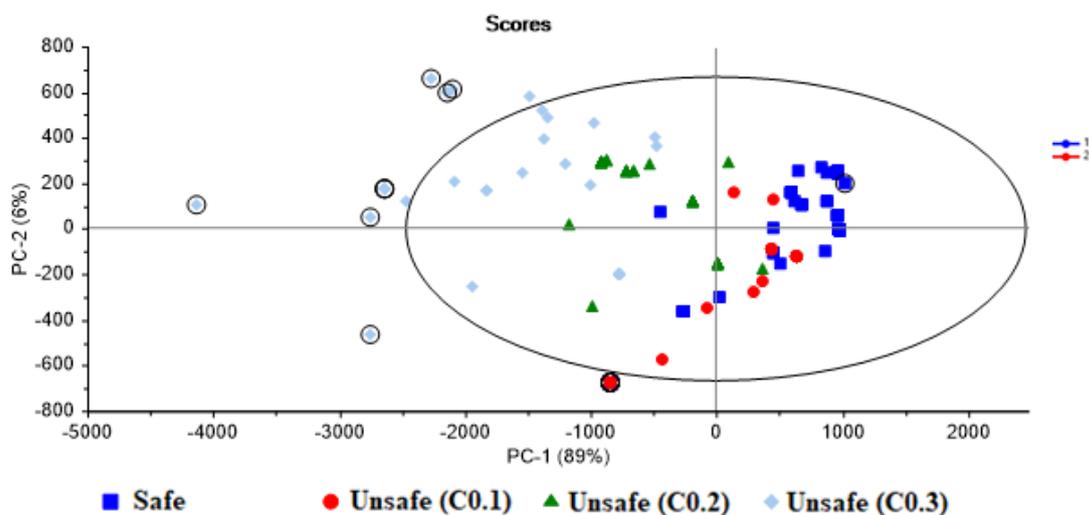


Figure 2: The loading Plot of the first principal component

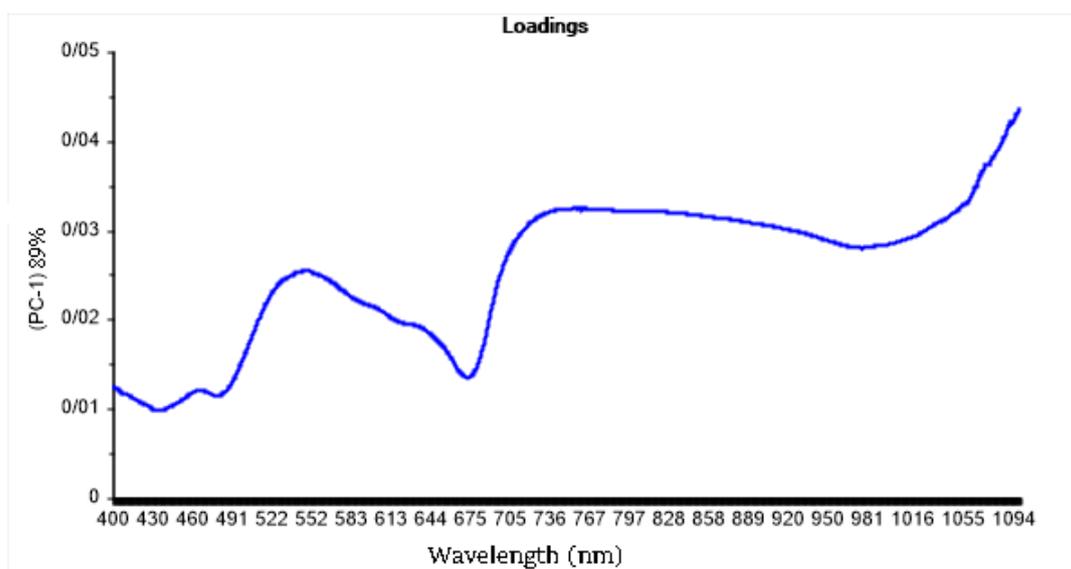
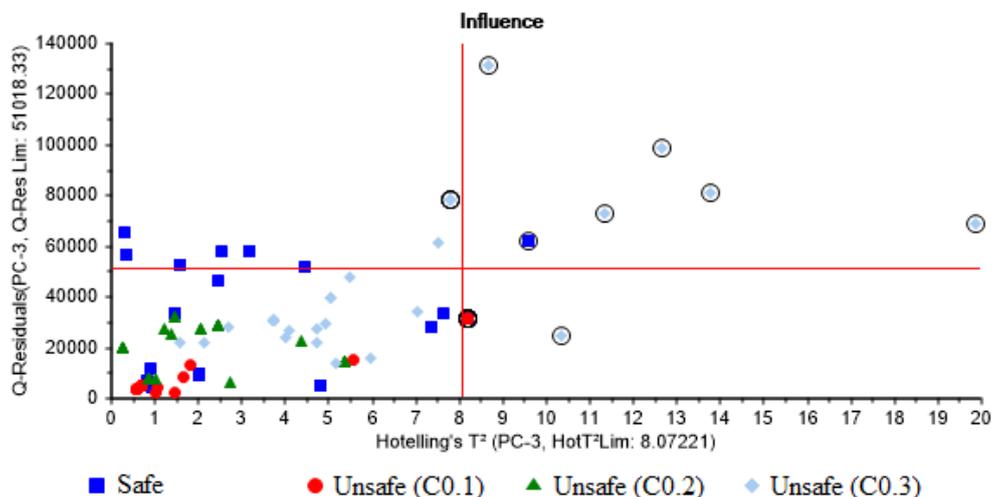


Figure 3: The Plot of influence from the lettuce samples (The samples with a circle around them are outliers).



### 3.2. PLS-DA Model

Partial least squares-discriminant analysis (PLS-DA) is known as a supervised qualitative method which can be applied for projecting modelling as well as for classification of variables [38, 39]. Results of PLS-DA model makes the same outcomes of Euclidean distance to centroids (EDC) which is as a classic method, and also it could be used as the method of linear discriminant analysis (LDA) (Gómez et al, 2006). In this paper, we used the PLS- DA to classify lettuce samples in more than one group which are safe and unsafe groups. After running PCA as a primary exploratory investigation of spectra, the models were tested by splitting the total 191 of samples into a training and test sets.

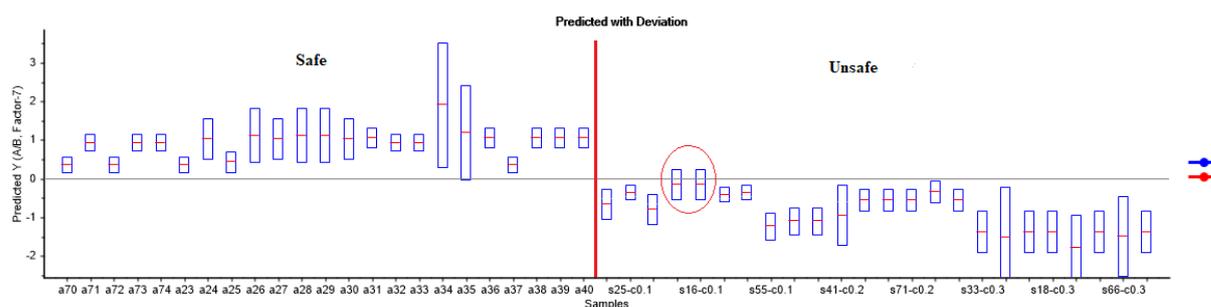
The statistical parameters obtained from PLS-DA models with various pre-processing spectra techniques to explore the lettuce samples with the presence of microbial contamination coming from contaminated lettuces with *E. coli*, are shown in Table 1. Some of the established models of calibration had not a great potential to classify the lettuces into two different classes. The best calibration models is developed with standard normal variate and second derivative (SNV+D2), as well as standard normal variate and first derivative (SNV+D1) with lowest standard error of cross-validation (SECV = 0.257) and highest coefficient of determination ( $r^2 = 93\%$ ). After using cross-validation method for the calibration set, 100% accuracy was found with a perfect classification. All reference lettuce samples were classified correctly for developed model using the SNV+D2 pre-processing technique. In calibration set, there were 23 safe group samples which all of them were correctly classified. In addition, in 24 samples of unsafe group only two samples are incorrectly classified. In consequence, the overall proportion of sample classification of unsafe group was 91.6%.

Table 1: The statistical parameters obtained from PLS-DA models various pre-processing spectra methods.

Preprocessing methods	LVs	Calibration set		Prediction set		
		R <sup>2</sup>	SECV	Total	safe	unsafe
MSC	7	70%	0.567	86.11	100	72
SNV	3	72%	0.537	86.11	100	72
D1	7	67%	0.582	80.55	100	62
D2	5	72%	0.525	86.11	100	72
MSC+D1	5	83%	0.416	86.11	100	73
MSC+D2	6	81%	0.397	86.11	100	73
SNV+D1	7	93%	0.269	95.8	100	91.6
SNV+D2	7	93%	0.257	95.8	100	91.6

Predicted outputs of the best model of calibration with measuring projected deviation are shown in Figure 4. In unscramble, there were 2 groups of the samples in prediction. The first one which contains 23 samples and the second one contains 24 of unknown samples. All samples of first group have predicted values close to 1 classifying these as belonging to group “safe”. The second-group samples have a predicted value around -1 which assigns them to group “unsafe”. For sample of s16-c0.1 (sample sixteen with 0.1 ml of microbial concentration) and s13-c0.1 (sample thirteen with 0.1 ml of microbial concentration), the predictions are close to 0, and have high uncertainties. It could be concluded that these samples cannot belong to any of the distinct groups, because of the measured projected deviation (uncertainty) are close the prediction value includes 0 in the plot.

Figure 4: Predicted outputs of the best model of calibration with measuring projected deviation (the projected response for the new samples shown as a horizontal red line. The blue box around the projected value spans the deviation in both directions and is an estimate of the projection uncertainty).



The results showed the possibility using Vis/NIR spectroscopy in the wavelength range from 450 to 994 nm for detection and evaluation of *E. coli* ATCC 8739 contamination in lettuce. As we see, there was close relation between safe samples and unsafe samples with 0.1 ml of *E. coli* solution. Therefore, the ability of Vis/NIR spectroscopy to definitive detection of *E. coli* contamination was conformed in those lettuces which have contamination above the 0.1 ml concentration.

Besides, PLS-DA, achieved good performance for nondestructive detection of contamination in intact lettuces that contain three different levels of the *E. coli* concentration. The decrease and increase of the microbial population amount in lettuce samples cause a complex process in lettuce and it effect on chemical components of lettuce. Therefore, the relationship between the changes of chemical components and the different *E. coli* content are very complicated and are most likely to demonstrate a non-linear relationship.

#### 4. Conclusion

In summary, the results of the current investigate proved using Vis/NIR spectroscopy in the wavelength range from 450 to 990 nm to rapid and accurate detection of *E. coli* ATCC 8739 affected in fresh lettuce leaves.

Classification the intact lettuce leaves spoiled with *E. coli* loads in two groups of safe and unsafe was performed by PLS-DA in different pre-processing methods. The optimum results achieved a Root Mean Squared Error Calibration (RMSEC) = 0.257% and correlation coefficient ( $r_c$ ) = 0.93 with standard normal variation (SNV) and second derivate (D2) pre-processing methods. It could be concluded that there is good investigative accuracy in detection of *E. coli* with minimal errors at a concentration  $>0.1$  mL in fresh lettuce leaf. Therefore, the Vis/NIR spectroscopy method could be used for non-destructive detection of the lettuces with regard to contaminate by *E. coli* in online applications of food safety. Further, the results were recommended using Vis/NIR spectroscopy technique as a fast and simple way to detect *E. coli* in other types of vegetable for further identifying. This shows that the NIR range was unable to detect the cell present in the samples at the early growth stage. It could be concluded When *E. coli* ATCC 8739 cell concentration reached bigger than 6.67 log CFU/mL, unsafe samples become statistically classified from the safe samples. Which may be attributed to the cell components or metabolic products.

#### References

- [1] Aske, N., Kallevik, H., and Sjoblom, J. (2001). "Determination of saturate, aromatic, 445 resion, and asphaltenic (SARA) components in crude oils by means of infrared and near-infrared spectroscopy," *Energy & Fuels*, Vol. 15, pp. 1304-1312.
- [2] André, J., and Lawler, I, R. (2003). "Near infrared spectroscopy as a rapid and inexpensive means of dietary analysis for a marine herbivore," *Marine ecology progress series*, Vol. 257, pp. 259 –266.

- [3] Elmasry, G., and Nakauchi, S. (2016). "Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality - a comprehensive review," *Biosystem engineering*, Vol. 142, pp. 53-82.
- [4] Elmasry, G., Barbin, D. F., Sun, D. W., and Allen, P. (2012). "Meat quality evaluation by hyperspectral imaging technique: An overview," *Critical Reviews in Food Science and Food Nutrition*, Vol. 52, pp. 689-711.
- [5] Foley, W. J., A. McIlwee, I. Lawler, L. Aragonés, A. P. Woolnough, and N. Berdig. (1998). "Ecological applications of near infrared reflectance spectroscopy: A tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance," *Oecologia*, Vol. 116, pp. 293–305.
- [6] Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., and Frias, J.M.. (2007). "Hyperspectral imaging e an emerging process analytical tool for food quality and safety control," *Trends Food Sci. Technol*, Vol. 18, pp. 590-598.
- [7] Gómez, A. H., He, Y., and Pereira A, G. (2006). "Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques," *Journal of food engineering*, Vol. 77(2), pp. 313-319.
- [8] Jamshidi, B., Mohajerani, E., and Jamshidi, J. (2016). "Developing a Vis/NIR spectroscopic system for fast and non-destructive pesticide residue monitoring in agricultural product," *Measurment*, Vol. 89, pp. 1-6.
- [9] Jamshidi, B., Mohajerani, E., Jamshidi, J., Minaei, S., and Sharifi, A. (2015). "Non-destructive detection of pesticide residues in cucumber using visible/near-infrared spectroscopy," *Food Additives & Contaminants*, Vol. 6, pp. 857–863.
- [10] Lu, Z, Y., Zhang, Q, L., and Liu, X. (2013). "New Magnetic Particle Cassette NDT Intelligent Detection Device," *Fourth International Conference on Intelligent Systems Design and Engineering Applications*, Zhangjiajie, Hunan, China.
- [11] Nicolai, B.M., Defraeye, T., De Ketelaere, B., Herremans, E., Hertog, M.L., Saeys, W., Torricelli, A., Vandendriessche, T., and Verboven P. (2014). "Annual Review of Food Science and Technology," Vol. 5, pp. 285-312.
- [12] Panagou, E. Z., Papadopoulou, O., Carstensen, J. M., and Nychas, G.J.E. (2014). "Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage," *International journal of food microbiology*, Vol. 174, pp. 1-11.
- [13] Park, B., Yoon, S.-C., Windham, W., R., Lawrence, K. C., Kim, M. S., and Chao, K.-L. (2011). "Line-scan hyperspectral imaging for real-time in-line poultry fecal detection," *Sensing and instrumentation for food quality and safety*, Vol. 5, pp. 25-32.
- [14] Suthiluk, P, Saranwong, S, Kawano, S, Numthuan, S, and Satake, T. (2008). "Possibility of using near infrared spectroscopy for evaluation of bacterial contamination in shredded cabbage," *Int J Food Sci Technol*, Vol. 43, pp.160-165.
- [15] Stewart, S., Priore, R.J., Nelson, M.P., Treado, P.J. (2012). "Raman imaging. Annu. Rev.," *Anal. Chem*, Vol. 5, pp. 337-360.

- [16] Williams, P., and Norris, K. (2001). "Near-infrared technology in the agricultural and food industries," Edited by St. Paul, Minn.: American Assn. of Cereal Chemists. pp. 135-137.
- [17] Wu, D., Sun, D.W. (2013). "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review d Part I: Fundamentals. Innov," *Food Sci. emerg. Technol*, Vol. 19, pp. 1-14.
- [18] Wang, Z., Ding, H., Lu, G., and Bi, X. (2014). "Use of a mechanical iris-based fiber optic probe for spatially offset Raman spectroscopy," *Opt. let*, Vol. 39 (13), pp. 3790-3793.
- [19] Wang, J, Kim, K. H, Kim, S, Kim, Y. S, Li, Q. X, and Jun, S. (2010). "Simple quantitative analysis of Escherichia coli K-12 internalized in baby spinach using Fourier transform infrared spectroscopy," *Int J Food Microbiol*, Vol. 144, pp.147-151.
- [20] Zhang, P., Wittmann, F. H., Zhao, T. J., Lehmann, E. H., Tian, L., and Vontobel, P. (2010). "Observation and quantification of water penetration into Strain Hardening Cement-based Composites (SHCC) with multiple cracks by means of neutron radiography," *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. 620 (2-3), pp. 414-420.
- [21] Zoubir, A. (2012). *Raman Imaging: Techniques and Applications*. Springer, New York, NY, USA. pp. 365-368.