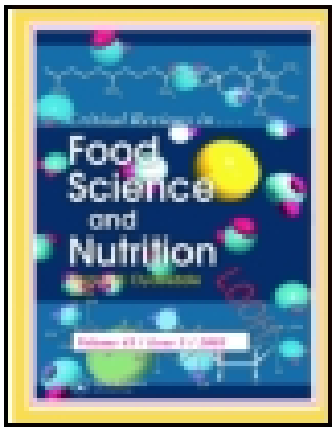

Critical Reviews in Food Science and Nutrition
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/bfsn20

Advances in Feature Selection Methods for Hyperspectral Image Processing in Food Industry Applications: A Review

Qiong Dai [a], Jun-Hu Cheng [a], Da-Wen Sun [a b] & Xin-An Zeng [a]

[a] College of Light Industry and Food Sciences, South China University of Technology, Guangzhou, 510641, China

[b] Food Refrigeration and Computerized Food Technology, Agriculture and Food Science Centre, University College Dublin, National University of Ireland, Belfield, Dublin 4, Ireland

Accepted author version posted online: 01 Apr 2014.

To cite this article: Qiong Dai, Jun-Hu Cheng, Da-Wen Sun & Xin-An Zeng (2014): Advances in Feature Selection Methods for Hyperspectral Image Processing in Food Industry Applications: A Review, Critical Reviews in Food Science and Nutrition, DOI: 10.1080/10408398.2013.871692

To link to this article: http://dx.doi.org/10.1080/10408398.2013.871692

Advances in Feature Selection Methods for Hyperspectral Image Processing in Food Industry Applications: A Review

Qiong Dai[a], Jun-Hu Cheng[a], Da-Wen Sun[a, b*], Xin-An Zeng[a]

[a] College of Light Industry and Food Sciences, South China University of Technology, Guangzhou 510641, China

[b] Food Refrigeration and Computerized Food Technology, Agriculture and Food Science Centre, University College Dublin, National University of Ireland, Belfield, Dublin 4, Ireland

* Corresponding author. Tel: +353-1-7167342, Fax: +353-1-7167493, E-mail: dawen.sun@ucd.ie, Website: www.ucd.ie/refrig; www.ucd.ie/sun.

Abstract: There is an increased interest in the applications of hyperspectral imaging (HSI) for assessing food quality, safety and authenticity. HSI provides abundance of spatial and spectral information from foods by combining both spectroscopy and imaging, resulting in hundreds of contiguous wavebands for each spatial position of food samples, also known as the curse of dimensionality. It is desirable to employ feature selection algorithms for decreasing computation burden and increasing predicting accuracy, which are especially relevant in the development of on-line applications. Recently, a variety of feature selection algorithms have been proposed that can be categorized into three groups based on the searching strategy namely complete search, heuristic search and random search. This review mainly introduced the fundamental of each algorithm, illustrated its applications in hyperspectral data analysis in the food field, and discussed the advantages and disadvantages of thses algorithms. It is hoped that this review should provide a guideline for feature selections and data processing in the future development of hyperspectral imaging technique in foods.

Keywords: Feature selection, spectral feature, texture feature, complete search, heuristic search, random search

1. Introduction

Hyperspectral imaging (HSI) is an advanced technique that integrates traditional spectroscopy and imaging to obtain both spectral and spatial information from an object. Over the past decade, as a non-destructive detection technique, hyperspectral imaging enjoyed a rapid development and wide usage in the researches of identifying both chemical and physical properties of food. It has been successfully applied in fruits for predicting color (ElMasry et al., 2007), constitution (Li et al., 2011), pesticide residues (Rajkumar et al., 2012), damage detection (Xing et al., 2005), tenderness (Naganathan et al., 2008), pH value, drip loss (Qiao et al., 2007a) and classification (Qiao et al., 2007b) in both beef and pork, and quality and freshness assessment in fish (Sone et al., 2012; Wu & Sun, 2013a, 2013b; Wu et al., 2012a; Cheng et al., 2013). As hyperspectral images are composed by hundreds of contiguous wavebands for each pixel of the object, the problems of such huge number of data increase the data processing load, resulting in the expenses of instrument and the complexity of online application. In addition, due to the negative influence caused by high colinearity and data redundancy among whole HSI dataset, the accuracy and robustness of classification or predictive models might be decreased (Burger & Gowen, 2011). Therefore, it is a very important step to select a small set of features that carry the most information of samples.

There are two aspects of feature selection in HSI, namely spectral features and image features. As only a few studies have analyzed the spatial features of hyperspectral images, the selection

methods of image features are relatively more hysteresis than that of spectral features. For spectral features selections, also known as wavelengths selection, selected wavelengths with sharp peaks or great statistical differences among samples were reported in some early studies. However, the bond vibrations of biological components in food are very complicated, leading to a common presence of broad envelope with few sharp peaks in spectrum, which makes it difficult to identify the key wavelengths directly as the number of samples increases (Gowen et al., 2007). Besides, it is impossible to select image features (color, shape, texture) in such a naked-eye way. Thus, advanced algorithms for variable selection should be applied to solve this problem.

The definition of feature selections should be introduced before further discussion of feature selection algorithms. According to Dash and Liu (1997), feature selection tries to find the minimally sized subset of features on the basis of the following two criteria: the accuracy of classification and prediction cannot significantly decrease; and the distribution of resulting class is as close as possible to that of original class. Generally speaking, the following four factors need to be considered in feature selection algorithms (Fig.1): (1) the search starting point to determine where to start searching and the search direction to determine the sequence of candidate subset; (2) searching strategy to find the optimal subset from searching space; (3) an evaluation function to evaluate the subset under examination; and (4) stop criterion to decide when to stop. The procedure of selecting the most informative variables in HSI dataset can be

defined as a problem of searching an optimized subset. Recently, a variety of feature selection algorithms have been proposed to HSI for food analysis that can be categorized into three groups based on the searching strategy (Liu et al., 2005). Group 1 guarantees to find the optimal subset of variables according to the evaluation criterion used, i.e., complete search (such as branch and bound), Group 2 obtains a relatively optimal subset by constantly adding or removing features from current feature subset under some certain order, namely heuristic search (such as stepwise regression), and Group 3 is based on a random search strategy combined with some certain intelligence, i.e., random search (such as genetic algorithms). Principal component analysis, regression coefficients, uninformative variable elimination, stepwise discrimination, successive projections algorithm, artificial neural networks, genetic algorithms, and competitive adaptive reweighted sampling have been employed to select the optimal subset of variables. All these studies suggest that a better predictive ability can be obtained while using the selected features (both spectral and spatial features) rather than the full feature space. Despite the good performances, these algorithms are far from perfect due to the presence of specific disadvantages. However, it should be noticed that progresses have been made among feature selection algorithms. Several main feature selection algorithms and their applications in HSI in the food field are described and discussed below and their relevant detailed information is illustrated in Table 1 and Table 2.

2. Spectral Feature Selection Methodologies and Applications

2.1. Complete Search

Complete search is a method that searches the whole results space in order to find the optimal subset according to the evaluation function used. There are two major selection methodologies namely exhaustive method and brand and bound method. The low efficiency of exhaustive method, which caused by the huge searching space (O (2^N)), make it unavailable for practical application. However, a search does not have to be exhaustive for the purpose of guaranteeing the optimal and the search space can be greatly reduced by applying different heuristic functions (such as branch and bound methods). Therefore, only a few subsets from the original search space are evaluated. The procedure of back track is commonly used to guarantee the optimality of the feature subset. Backtracking can be implemented by various techniques such as branch and bound, beam search, and best first search (Jain & Zongker, 1997).

2.1.1. Branch and bound

Branch and bound (BB or B&B) is a general algorithm for finding the optimal subset of features (Narendra, & Fukunaga, 1997). It is composed of a systematic enumeration of all possible solutions (features), in which large subsets of fruitless solutions need to be discarded, then the optimized results are obtained by using upper and lower estimated bounds of the quantity. Three main steps of branch and bound algorithm are described as follows (Nakariyakul

& Casasent, 2004).

The first step is the splitting procedure (called branching), which divided the candidate solution space (S) into two or more smaller subsets (S_1, S_2, …, S_n), the second step is computing procedure (called bounding), which computes upper and lower bounds for the value of the solution within each subset of S (S_1, S_2, …, S_n), and the third step (pruning step) is the main idea, which indicates that if the lower bound for some tree node (set of candidates or features) A is greater than the upper bound for some other node B, then A may be free from the further search. The recursion stops until only a single element left in the current candidate set S or the upper bound for set S matches the lower bound. Either way, the element of S will be the optimal subset of candidates or features.

A few applications of complete search methodologies are available for processing hyperspectral imaging data in foods. For example, Nakariyakul and Casasent (2007) employed a modified branch and bound (MBB) algorithm to select a small number of wavelengths for detecting skin tumors on chicken carcasses. The modified algorithm used a sub-optimal algorithm (forward selection) to rank all the wavelengths from best to worst, and tree was then constructed according to this ranking order. As a result of calculating MBB, only 5 wavelengths (510, 560, 580, 690, and 730 nm) out of sixty wavelengths were identified as the optimal wavelengths that contained the most information of spectral reflectivity between normal and tumor skins of chicken carcasses, leading to the reduction of 91.7% in computing load with only

2 false alarms.

2.1.2. Advantages and disadvantages

As indicated previously, complete search strategy can find the optimal subset of variables from the full spectrum, and this is its most important advantage. However, by searching through the whole results space, it would produce problems of time-exhausting as the number of features increase, especially for the case of hyperspectral data. To some extent, B&B or MBB can reduce the searching time of optimal wavelengths by pruning branches. From the perspective of statistics, such a reduction is not enough for large number of branches and nodes because its magnitude of searching time does not differ much from that of exhausted methods. Another limitation of B&B is that, the efficiency and accuracy of such an algorithm greatly depends on the prioritization of decisions and the accuracy of estimation function at each node. Therefore, the complete search might be not suitable for variable selections of hyperspectral data due to the latent defect presented in Table 3.

2.2. Heuristic Search

Heuristic search obtains a relatively optimal solution by adding or removing features under some certain orders such as forward selection, backward elimination, and bidirectional selection (Liu et al., 2005). In each iteration of heuristic search, all remaining features yet to be selected

(rejected) are considered for selection (rejection). There are many variations to this simple process, but the generation of subsets is either by increasing or decreasing incremental. Compared to complete search, the efficiency of heuristic search is greatly improved by searching a small space (O (N^2) or less). Heuristic search based upon wavelength selecting algorithms for analyzing hyperspectral image data in food includes principal component analysis, spectral derivatives, regression coefficients, stepwise regression, uninformative variable elimination, successive projections algorithm, and artificial neural network.

2.2.1. Principal component analysis and spectral derivatives

Both principal component analysis (PCA) and spectral derivatives are feature selection methods by observing visually spectral profile and conducting manual selection. Specially, principal component analysis, an eigenvector-based algorithm, is commonly used as an exploratory method for feature selection in hyperspectral images. PCA may be understood as maximizing the variance of the projection coordinates. In PCA, the original data with high collinearity can be rapidly concentrated into a smaller set of principal component score images (PCs) that are linear transformations of all original variables. The fundamentals of PCA are described below (Abdi & Williams, 2010).

The main goal of PCA is to find a group of orthogonal basis for better expressing the original spectral data matrix X . Then PCA breaks apart X into a principal component matrix called the

eigenvectors or loadings (F), in which the dimensions are orthogonal to each other and with maximum variations, and the corresponding scaling coefficient matrix called the scores (S) as the following equation:

$$X = SF + e_a \quad (1)$$

where X is a $n \times k$ matrix of spectral data, S is a $n \times f$ matrix of score values for all of the spectra, and F is a $f \times k$ matrix of eigenvectors. e_a is the residual spectra matrix, while n is the number of samples, k is the number of wavelengths, and f is the number of principal components.

The score values in S determine the importance of original variables in the principal components. Therefore, the variables having high score values can be considered as the effective wavelengths that represent the most variance and contribution of dataset.

Numerous applications of PCA for HSI band selection in foods have been widely presented in literature. Currently, these applications have been focused on bruises detection of fruits such as apples (Xing et al., 2005; Xing & De Baerdemaeker, 2005) and oranges (Li et al., 2011), internal defects and surface infections of vegetables such as pickles (Ariana & Lu, 2010) and onions (Wang et al., 2012), quality classification of meat such as pork (Qiao et al., 2007a), lamb (Kamruzzaman et al., 2011) and turkey hams (ElMasry et al., 2011a), and diseased tissue detection of poultry such as chicken (Chao et al., 2002). In all these studies, a high accuracy for

classification was obtained while using only a small set of wavelength bands rather than the full spectrum, which proved that PCA was a useful method for selecting the most informative wavelengths. The general process of PCA for wavelength selection is described as follows, the full spectrum extracted from the interest region of all samples were first analyzed by PCA, then variables at peaks and valleys (with high loading values) were selected as effective wavelengths by observing the spectral profile from the first several PCs. For example, ElMasry et al. (2011a) analyzed the eigenvectors of first three PCs that extracted from the hyperspectral image of turkey ham samples. The peak wavelengths of the first PC was mainly centered at 1061, 1174, 1215, 1325, 1436, and 1641 nm while only two peak wavelengths (980 and 1141 nm) were found in the second PC. As a result, eight wavelengths namely 980, 1061, 1141, 1174, 1215, 1325, 1436, and 1641 nm were identified as the important wavelengths for ham quality classification. In addition, Li et al. (2011) chose the second PC (691 and 769 nm) and the third PC (630, 691, 769, 786, 810, and 875 nm) to detect the common defects on orange, leading to good accuracies of 91.5% and 93.7% with no false positives based on both sets of selected wavelengths in the visible (Vis) and near infrared (NIR) regions. On the other hand, in the application of HSI for lamb muscle discriminations, five wavelengths (974, 1074, 1141, 1211, and 1308 nm) from PC1 and one wavelength (934 nm) from PC4 were selected as the optimum wavelengths that explained most variance among lamb muscles with a perfect accuracy of 100% for discrimination (Kamruzzaman et al., 2011).

The first and second derivatives are two commonly used derivative methods of hyperspectral wavelength selection. The bandwidth of each wavelength can be considered as a variable of hyperspectral image. Hence, the first and second derivative methods are carried out to explore the bandwidth variables for added information. Generally speaking, the adjacent bands that differ a lot should be reserved for the prediction of food attributes, while similar adjacent bands should be deleted. The first derivative measures the difference of two adjacent bands while the second derivative measures the linear deviation of three adjacent bands. Thus in the first derivative, if two adjacent bands do not differ much, then the food attributes can be characterized with only one band. On the contrary, the second derivative identifies if two adjacent bands can linearly interpolate the third band, then the third band can be removed. The mathematical descriptions of the first and second spectral derivatives are illustrated by the following two equations, respectively (Bajwa et al., 2004).

$$D_1 = \frac{\partial I(x, y, \lambda)}{\partial \lambda} \qquad (2)$$

$$D_2 = \frac{\partial^2 I(x, y, \lambda)}{\partial \lambda^2} \qquad (3)$$

where I is the spectral reflectance value, x, y are spatial locations, and λ is the wavelength band. Thus, the larger the D_1 and D_2 are, the higher the information value of the band is.

The second derivative has been more frequently used than the first derivative in recent studies of hyperspectral imaging application in foods. This might be because that inverse peaks appeared

at the location of the original ones for the case of second derivative rather than the first derivative. The second derivative by Savitzky-Golay algorithm is a common choice in applications of HSI in the food field such as classification of pork (Barbin et al., 2012a; Barbin et al., 2012b), red meat (Kamruzzaman et al., 2012a), prediction of total viable count (TVC) and psychrotrophic plate count (PPC) on pork (Barbin et al., 2013a), and determination of *Enterobacteriaceae* on chicken (Feng et al., 2013). The second derivative turns the smooth raw spectral into a kind of sharper plots, in which there are more non-overlapping peaks, more essential spectral details and more deviation of spectral bands. It is very helpful in spectral interpretation and signal intensification by maintaining the most important wavelengths and cutting down the most redundant bands. However, the second derivative is very sensitive to noise, proper filtering is thus necessary before calculation of derivatives. Therefore, Savitzky-Golay algorithm is used for this purpose. Specially, the reflectance spectrum is filtered by Savitzky-Golay filter with different sizes according to the noise level of raw spectrum, which minimizes the negative effects of noise in the following derivatives. At last, the high peaks and low valleys at the second derivative plots gained by Savitzky-Golay are considered as the most important wavelength bands. For example, Barbin et al. (2012a) adapted the second derivative by Savitzky-Golay method to identify six key wavelengths (960, 1074, 1124, 1147, 1207, and 1341 nm) for pork classification using hyperspectral imaging. In his later research of predicting microbial contamination in porcine meat (Barbin et al., 2013a), the second derivative by

Savitzky-Golay method with nine point interval and second order polynomial filters were used to identify seven wavelengths (964, 1128, 1151, 1301, 1341, 1395, and 1635 nm) from full spectrum that were most related to the spoilage. Additionally, as a result of the second derivative by Savitzky-Golay method, Kamruzzaman et al. (2012a) chose six wavelengths (957, 1071, 1121, 1144, 1368, and 1394 nm) as the most effective wavelengths in discrimination of red meat.

2.2.2. Regression coefficients, uninformative variable elimination, and stepwise regression

Regression coefficients, uninformative variable elimination (UVE) and stepwise regression algorithms (SPA) are all based on analysis of coefficients of regression models. Regression coefficient method, also called β-coefficients method, uses the regression coefficients resulting from the calibration model (PLSR) to choose the effective wavelengths (Chong & Jun, 2005). For a better understanding of regression coefficients, it is necessary to introduce PLS regression algorithm firstly (Geladi & Kowalski, 1986).

The PLS algorithm defines a set of orthogonal projection axes W from spectral matrix X (n × m), called PLS-weights, the wavelength scores T and their corresponding linear relations as described below:

$$X = T \times P^T + E; \quad (4)$$

$$y = T \times q^T + f; \quad (5)$$

$$T = X \times W^*; \quad (6)$$

$$W^* = W \times \left(p^T W \right)^{-1} \quad (7)$$

Then the regression coefficients (β) are calculated from: $\beta = W^* q^T = W \left(P^T W \right)^{-1} q^T$

Finally, the PLS model can be expressed as a multiple linear regression: $y' = XW_K \beta = T\beta$

where β is the vector containing the regression coefficients (m$\times$1) for each variable obtained from the PLS model, y is the reference value of the attribute of interest (n$\times$1), q is the y-loadings (1$\times$k), T is the latent variable space (n$\times$k), W^* is the m$\times$k matrix of X-weights, P is the m$\times$k matrix of X-loadings, and y' is the corresponding predicted value of the attribute of interest (n$\times$1).

Generally speaking, the wavelengths that with the high absolute values of β-coefficients could be selected as the feature wavelengths whereas the ones with the low absolute values of β-coefficients could be completely removed due to their little contribution in prediction.

Among the large family of dimensionality reduction algorithms, regression coefficient is one of the most frequently employed techniques for wavelength selections in hyperspectral images of foods. In contrast with PCA, regression coefficients method is often used in the wavelength selections for quantitative prediction of food attributes. These applications involve the prediction of water, fat and protein contents, the color, pH, tenderness, water holding capacity, sensory assessment of beef (ElMasry et al., 2011b, 2012), pork (Qiao et al., 2007; Barbin et al., 2012b, 2013a), lamb (Kamruzzaman et al., 1012c, 2013a), hams (Talens et al., 2013), the moisture content, total soluble solids, acidity of strawberry (ElMasry et al., 2007), the firmness of banana

(Rajkumar et al., 2012), the macronutrients content of oilseed (Zhang et al., 2013), the *Enterobacteriaceae* of poultry (Feng et al., 2013), and the adulteration of lamb meat (Kamruzzamanet al., 2013b). Also there is an application for classification of flesh salmon stored under different atmospheres (Sone et al., 2012). In all these applications, the effective wavelengths selected from regression coefficients method performed equally or more efficient than full wavelengths. Specifically, once calibration models based on the full spectrum was established, the regression coefficients at each wavelength of the PLSR model could be plotted. Then only the wavelengths with the highest absolute values of regression coefficients at the peaks or valleys could be considered as the effective wavelengths. The final set of wavelengths was obtained after removing the wavelengths with low signal noise ratio. Kamruzzaman et al. (2012c) selected six wavelengths (960, 1057, 1131, 1211, 1308, and 1394 nm) with high weighted regression coefficients resulting from PLSR models for both water and fat contents of lamb meat (R_P^2 of 0.84, 0.87, respectively). Regression coefficients method was also used for the wavelength selection of HSI in banana quality prediction by Rajkumar et al. (2012), who found eight wavelengths (440, 525, 633, 672, 709, 760, 925, and 984 nm) for total soluble solids, moisture, and firmness prediction of banana, leading to a good R_P^2 of 0.85, 0.87, and 0.91, respectively. Good results were also achieved (R_P^2 of 0.87) by using three wavelengths (930, 1121, and 1345 nm) selected from weighted PLS regression coefficients for predicting *Enterobacteriaceae* loads in chickens by Feng et al. (2013).

UVE is a method of wavelength selection method based on regression coefficients of PLS regression model. It eliminates wavelengths that provide no or little information to the established regression model by setting a threshold. The main steps of UVE are summarized as follows (Centner et al., 1996):

Step 1: A PLS regression model is established based on spectral data X and their corresponding reference parameter value y,

Step 2: The regression coefficient of PLS model for each wavelength is calculated, and

Step 3: The reliability of each wavelength is analyzed by the following formula:

$$C_i = \frac{mean(b_i)}{s(b_i)} \qquad (8)$$

where mean(b_i) and s(b_i) are the mean value and standard deviation of the regression coefficient (b) of wavelength (i) from PLS, respectively. The larger the reliability C is, the more important the corresponding wavelength becomes. Once a cut-off value (threshold) is established, the variable with reliability below the threshold will be eliminated, and the higher ones will be retained.

Advantages of UVE include its simple procedure, fast execution, convenience, and high accuracy. UVE can eliminate the variables that have no more informative variables for modeling than noise. Besides, employing variables selected by UVE to establish models can avoid the risk of over-fitting and improve the accuracy of predicting. However, UVE is not suitable for selecting wavelengths directly because the number of wavelength that UVE gets is still large.

Therefore, a further selection for the most important variables is critical. As shown in several studies (Di et al., 2009; Wang et al., 2012), the generalization ability and stability of predicting model can be greatly improved by using fusion method of wavelength selection algorithms such as UVE-SAP, and UVE-SPA. For instance, Wang et al. (2012) applied UVE-SAP to identify the effective wavelengths for predicting apple firmness using hyperspectral scattering image. The UVE algorithm improved the prediction ability of the PLS model as it took into consideration the combination of spectral information with the actual firmness. However, the algorithm was unstable during the wavelength selection due to the extra noise matrix. On the contrary, the supervised affinity propagation (SAP) algorithm was an efficient and fast clustering algorithm, but it lacked supervision over the effective parameter values (actual firmness) and focused on analyzing the similarity among wavelengths. A better performance was obtained by using only one thirds of the full wavelengths. Another research for classification of prawns employing UVE-SPA to carry out the task of important wavelengths selection was conducted by Di et al. (2009). As a result of UVE-SPA calculation, twelve wavelengths (392, 431, 517,551, 595, 627, 676, 734, 760, 861, 943, and 1018 nm) were identified as the important wavelengths that carried the most information for the classification.

Stepwise regression is another regression model based method for variable selections. Practically, its step-by-step iterative construction of a regression model implements the automatic selection of independent variables. Stepwise regression can be achieved either by forward

methods or backward methods, or by combing both methods. The fundamentals are illustrated below (Xu & Zhang, 2001).

Given the standard linear model $y = \sum_{i=1}^{k} \beta_i X_i + e$ with the residuals *e*, assumed to be NID (0, δ^2). F_s-values are calculated for all variables after the inclusion of a new variable as follows:

$$F_s = \frac{\frac{SS_{k-p}}{k-p}}{\frac{SST - SSR_k}{n-k}} = \frac{\frac{SS_{k-p}}{k-p}}{MSE_k} \quad (9)$$

where SSR_k is the total regression sums of squares, SS_{k-p} is the additional regression sums of squares, SST is the total sums of squares (of y), MSE_k is the residual sums of squares, and k-p and n-k are the two freedoms of F distribution.

According to F_s-values calculated for all variables after the inclusion of a new variable, the variable with non-significant F_s statistic should be removed from the model. The procedure ends until F_s-value testing for inclusion of a variable becomes non-significant. This method allows for the identification of variables that have become redundant by subsequently adding new variables into the regression equation, which may overcome the shortcomings of the forward selection procedure.

Several attempts have been made to use stepwise regression for wavelength selection of HSI in meats such as beef (Cluff et al., 2008; Wu et al., 2012c), pork (Tao et al., 2012), and chicken (Feng & Sun, 2013b). For beef (Wu et al., 2012c), eight wavelengths (485, 524, 541, 645, 700,

720, 780 and 820 nm) for tenderness, seven (653, 678, 722, 868, 875, 920 and 1050 nm) for L*, seven (465, 575, 614, 635, 671, 724 and 978 nm) for a* and eight (486, 524, 540, 645, 700, 721, 780 and 954 nm) for b* were determined as the most important wavelengths for prediction model. Then these wavelengths were used as input of the calibration model and good performances were achieved with R_{cv} of 0.91, 0.96, 0.96, and 0.97 for tenderness, L*, a*, and b*, respectively. As for chicken (Feng & Sun, 2013b), five wavelengths (954, 957, 1138, 1148 and 1328 nm) were selected by stepwise regression for the simplified PLSR model to predict total viable count in raw chicken fillet. With the selected wavelengths, the simplified model gave better performance (R_{cv}=0.88) than the original wavelengths although the RMSECV were slightly increased to 0.69. For pork (Tao et al., 2012), six wavelengths (490, 559, 580, 553, 564 and 536 nm) were used to establish predicting model for pork tenderness and five wavelengths (525, 875, 858, 534 and 530 nm) were identified as the contributing wavelengths for predicting *E.coli* contamination of pork.

2.2.3 Successive projections algorithm and artificial neural network

The successive projections algorithm (SPA) is a novel variable selection algorithm. It uses the vector projection analysis for finding the minimum of redundant information contained variable group. At the same time, it improves the speed and rate of model by reducing the number of variables (Araújo et al., 2001).

The basic principle of SPA, as shown in Fig. 2, is that it firstly begins with one wavelength, and then incorporates another one at each iteration until a specially defined number (N) of wavelengths is finished. The aim of this method is to choose a set of wavelengths, which are the most representative, for solving collinearity problems. Several steps of SPA are needed as described below, assuming that the first wavelength *k* (0) and number *N* are known.

Step 0: Before the first iteration (*n*=1), let $x_j = j$ th column of X_{cal}; $j = 1,...,J$.

Step 1: Let S be the set of wavelengths, which have not been selected yet. That is, $S = \{ j$ such that $1 \le j \le J$ and $j \notin \{k(0),...,k(n-1)\}$.

Step 2: Calculate the projection of x_j on the subspace orthogonal to $\boldsymbol{x}_{k(n-1)}$ as

$$Px_j = x_j - \left(x_j^T x_{k(n-1)}\right) x_{k(n-1)} \left(x_{k(n-1)}^T\right) \left(x_{k(n-1)}^T x_{k(n-1)}\right)^{-1} \quad (8)$$

Step 3: Let $k(n) = \arg\left(\max \left\| Px_j \right\|, j \in S\right)$, for all $j \in S$, (9)

where *P* is the projection operator.

Step 4: Let $x_j = Px_j$, $j \in S$.

Step 5: Let $n = n+1$. If $n < N$, then go back to Step 1.

End: The resulting wavelengths are $\{k(n); n = 0,...,N-1\}$.

If N and k(0) are unknown in priori, then a method proposed by Araújo et al. (2001) is helpful to find the optimum values N^* and $k^*(0)$.

The SPA has been identified as an effective method for wavelengths selection in hyperspectral imaging systems by several authors (Wu et al., 2012b; 2012c; Kamruzzaman et al., 2013b; Wang

et al., 2012). Wu (2012b) attempted to use the SPA method for allocating the effective wavelengths related with the color components (L*, a*, b*) of salmon color. As a result, only four (1295, 1362, 1440 and 1527 nm), six (964, 1081, 1105, 1158, 1295 and 1406 nm), and ten (964, 1024, 1084, 1105, 1131, 1161, 1212, 1295, 1440 and 1530 nm) wavelengths were selected as the effective wavelengths for L*, a*, and b* prediction, respectively. Kamruzzaman et al. (2013b) used SPA method and selected eleven wavelengths (934, 964, 1017, 1081, 1144, 1215, 1265, 1341, 1455, 1615 and 1655 nm) to be the most contribution of tenderness variation. SPA was also conducted to choose the most important wavelengths for predicting dry matter content for potato (Zhou et al., 2012), moisture content of prawn (Wu et al., 2012c) and beef (Wang et al., 2012). With these key wavelengths selected by SPA as input of multiple linear regression (MLR), PLSR or partial least squares (LS) - support vector machines (SVM), good prediction results were obtained in these studies. Take an example of the moisture detection of prawn, the SPA-PLSR, SPA-LS-SVM and SPA-MLR based on twelve wavelengths (428, 445, 544, 569, 629, 672, 697, 760, 827, 917, 958 and 999 nm) gave a high of 0.955, 0.984 and 0.965 with a low RMSEP of 2.585, 1.502 and 2.304, respectively. There were similar performance between the models established based on the full spectrum (F-PLSR and F-LS-SVM) and the selected variables of SPA (SPA-PLSR, SPA-LS-SVM). However, a decreased RMSEP was found in both SPA-PLSR and SPA-LS-SVM models, which indicated that SPA could reduce the chance of overfitting and improve the accuracy of prediction.

Artificial neural network (ANN) has been widely used for machine learning and pattern recognition. Generally speaking, ANN consists of three layers: an input layer, an output layer and a hidden layer. The nodes of input layer represent the spectral responses of each wavelength from full spectrum. Usually, a small number of nodes on hidden layer might be set due to the great number of wavelengths. However, the volume of output layer depends on the practical needs for prediction or classification.

The importance of each variable (wavelength) for the ANN model can be calculated by the following equation (Vemuri, 1988).

$$M = \frac{\sum_{j=1}^{n_H}\left[\left(\frac{|I|_{p_j}}{\sum_{k=1}^{n_p}|I|_{p_{j,k}}}\right)|O|_j\right]}{\sum_{i=1}^{n_p}\left(\sum_{j=1}^{n_H}\left[\left(\frac{|I|_{p_{i,j}}}{\sum_{k=1}^{n_P}|I|_{P_{i,j,k}}}\right)|O|_j\right]\right)} \quad (10)$$

where M is the index measuring the importance of input variable, n_p is the number of input variables, n_H is the number of hidden layer nodes, $|I|_{pj}$ is the absolute value of the hidden layer weight corresponding to the *p*th input variable and the *j*th hidden layer, and $|I|_j$ is the absolute value of the output layer weight corresponding to the *j*th hidden layer. The index M value is calculated for each input node and then normalized into the 0-1 range. The higher the M value, the more important the node (wavelength) is for the classification or prediction model.

Only a few applications of ANN are available for identifying the important wavelengths for

HSI in foods. As the first attempt, ElMasry et al. (2009) applied ANN to choose the feature wavelengths for detecting chilling injury in apples. According to the calculated M values, 5 wavelengths (717, 751, 875, 960 and 980 nm) were chosen as the informative wavelengths, which were sensitive to any firmness changes that occurred in the apple during chilling injury. With the selected wavelengths instead of the full spectral range (826 wavelengths), the correlation coefficients between the actual and predicted firmness obtained from ANN models were 0.93, 0.91, 0.92 for training, testing and validation sets, respectively.

2.2.4. Advantages and disadvantages

In recent years, heuristic search has rapidly been developed and widely used for wavelengths selection of HSI. It has been proved that heuristic search methods are very effective in extracting important wavelengths that explain most the variation of samples. Compare to complete search, a common advantage of heuristic search is that they avoid the complete searching from the whole spectrum, leading to a relative smaller computing complexity. However, the reduction of searching time is at the expense of increasing the risk of losing optimal subsets. Besides, there are many other disadvantages among the heuristic search methods (Richard, 1990). Take PCA and spectral derivatives as example, drawbacks of such manual selection methods are that they could not extract optimal wavelengths subsets that most explain the variance of samples, instead some correlation wavelength or noise wavelength

might be selected. As for regression based methods, the accuracy of regression coefficient method is mostly based on subjective experience of threshold selection. The number of variables selected from UVE is still too large so that further selections are still required for modeling. Since the simple projection operations that SPA employs, the selected variables may have low signal noise ratio (SIN) or be useless for multivariate calibration, leading to negative effects on the precision of prediction or classification (Table 3).

2.3. Random Search

Random search is rather new in its use in feature selection methods compared to the other two categories. It starts with a randomly selected subset of variables and proceeds in two different ways. One is from heuristic search that injects randomness into the above classical heuristic approaches. The other is to generate the next subset in a completely random way, also called the Las Vegas algorithm (Karnopp, 1963). For both ways, the purpose of using randomness is to escape local optima in the search space and to avoid dependence of available resources. Although the search space is $O(2^N)$, random search methods typically only search a few number of subsets by setting a maximum times of possible iterations. However, some parameters should be set in advance for each random search method, which is an important task for achieving good results. As for GA, it needs an initial population size, a crossover rate and a mutation rate.

2.3.1. Genetic algorithms

Genetic algorithms (GA) assume that better results, achieved from experimental conditions, will prevail over the worst ones. At the same time, an improvement can be obtained by some sort of recombination together with some random changes. From this perspective, experimental conditions are considered as the genome, of which genes are the variables taking part in the process. Then the fitness of each experimental condition is measured by an optimized response. Generally speaking, GA made up of five elemental steps: (1) coding of all variables; (2) initiation of population; (3) evaluation of the responses; (4) reproductions; and (5) mutations. Steps 3-5 would be alternated until a termination criterion is achieved. The criterion can be either based on a lack of improvement in the response or simply on a maximum number of generations or on the total time allowed for the elaboration (Leardi & Gonzalez, 1998).

Due to the randomness nature of GA, the selected wavelengths might be different during different implementations. Therefore, it is necessary to execute GA programs repeatedly to determine the initial wavelength candidates, which are under the assumption that the common wavelengths selected by different runs of GA have great importance in accounting for the targets of interest. As a first attempt, Feng and Sun (2013b) ran five times of GA programs, then five wavelength sets were produced and the common wavelengths agreed by more than 3 times were chosen. As a result, the best model was finally established based on 14 wavelengths in five wavelength regions (1138-1155 nm, 1195-1198 nm, 1392-1395 nm, 1452-1455 and 1525-1529

nm), which achieved an R_p of 0.88 and RMSECV of 0.65 for predicting *Pseudomonas* loads in chicken fillets.

2.3.2. Competitive adaptive reweighted sampling

Another random search method for wavelength selection is competitive adaptive reweighted sampling (CARS), which is based on the simple but effective principle 'survival of the fittest' (Li et al., 2009). Suppose N as the number of wavelength subsets from N sampling run. Generally speaking, there are six steps in each sampling run of CARS: (1) k samples are selected by applying Monte Carlo strategy (Babai, 1979); (2) a calibration model, taking an example of PLS, is built using the selected k samples and the regression coefficient of each variable is produced; (3) exponentially decreasing function (EDF) is used to remove the wavelengths with relatively small absolute regression coefficients in a stepwise and efficient way; (4) adaptive reweighted sampling (ARS) is applied to realize a further competitive selection of wavelengths, which is similar to the 'survival of the fittest' principle in Darwin's Evolution Theory; (5) cross validation is utilized to evaluate the performance of each subset; and (6) the subset with the lowest RMSECV value is selected as the feature subset.

As a novel wavelength selection method, CARS received an increasing interest in selecting effective wavelengths in foods. Wu and Sun (2013a) applied CARS to select the most useful wavelengths for predicting the TVC level of salmon flesh. As a result of the CARS calculation,

eight wavelengths (495, 535, 550, 585, 625, 660,785, and 915 nm) were identified to explain the most TVC distribution of salmon flesh. The CARS-PLSR model established using the selected important wavelengths was considered to be the best model with R_p^2 of 0.985 and RPD of 5.127. In their later research for predicting water holding capacity of salmon (Wu & Sun, 2013b), good results of using CARS together with PLSR were obtained for the water-holding capacity determination of salmon. Only thirteen (430, 445, 450, 510, 605, 620, 760, 765, 830, 955, 965, 975 and 995 nm), twelve (440, 450, 520, 595, 600, 615, 765, 830, 885, 925, 975 and 995 nm), nine (420, 560, 680, 620, 625, 695, 700, 835 and 840 nm), and twelve (600, 605, 615, 755, 850, 860, 865, 890, 935, 940, 960 and 995 nm) wavelengths were identified as the important wavelengths, by CARS for the prediction of four water-holding capacity indices, respectively. New reduced model established based on selected wavelengths (CARS-PLSR, CARS-LS-SVM) resulted in a decrease of computing load by 89.3%, 90.1%, 92.6%, and 90.1% and a similar RPD for four water-holding indices, respectively (Wu & Sun, 2013b).

2.3.3. Advantages and disadvantages

As the newly developed wavelength selection methods, the random search algorithms (GA, CARS) were applied to the analysis of HSI data in foods recently. They usually produce a minimum subset of feature variables with no order among the selected features. Compared with the two above search strategies (complete search and heuristic search), the random search can

obtain an approximation solution of optimal subset of wavelengths rapidly by using some intelligent strategies. However, the presences of time-consuming and unstable results are the most limitations of the genetic algorithm. Specially, running time would increase when there are many feature wavelengths. Besides, different minimum subsets of features at each different execution may be produced, using the same training data and keeping the same input parameters. Therefore, researchers usually need to run a certain number of independent executions to obtain a feature subset, which represents the average of the minimum subsets (Table 3).

3. Image Feature Selection Methodologies for Hyperspectral Imaging Analysis

3.1 Image Features of Hyperspectral Image

As mentioned previously, the analyzed spectral information of hyperspectral image is one-dimensional vector. In other words, only the reflectance or absorption value of a pixel (mean spectra of ROI) was extracted to represent the sample, leading to the ignorance of the variation and relationship of the pixels among the image. Actually, the image contains a large amount of information such as shape, color and texture. Especially, texture feature plays a significant role in image analysis, which possesses important information about the structural arrangement of surfaces and their relationship to the surrounding environment. There are many ways to extract texture features. Practically, the gray-level co-occurrence matrix (GLCM) and wavelet analysis approach are two most popular statistical methods that are used to measure the textural

information of images (Haralick, 1973; Mallat, 1989). As a result of GLCM calculation, many textural features can be extracted from an image such as angular second moment, correlation, inertia, absolute value, inverse difference, entropy and maximum probability. Wavelet analysis is also a powerful signal analysis tool for characterization of visual texture of digital images at multiple resolutions. However, no matter which method is used, a large number of textural data can be produced. Besides, ordinary images only have three color channels, while hyperspectral images consist of hundreds of images. Therefore there is a huge number of texture features even for medium-sized hyperspectral image, and thus it is very necessary to conduct a selection for textural features.

There are two different ways of textural feature selection by using the feature selection algorithms. In the first approach, all hyperspectral image slices are subjected to extracting the texture features, followed by feature selection methods. In the second approach, feature selection methods are performed to select several informative images and texture features are then extracted only from these selected images. The second approach is more commonly used despite its low classification accuracy.

3.2 Texture Selection Methods for Hyperspectral Images of Foods

The potential ability of using feature selection in a high dimensional space has been successfully proved (Saeys et al., 2007). However, there is a lack in applying feature selection

algorithms to select texture features in hyperspectral images. Until recently, some attempts have been made to use PCA, β-regression coefficients, and stepwise regression for texture feature selection of HSI in the food field (Table 2).

PCA method has recently been used to select the texture features for quality classification of beef, lamb, pork meat and fish fillets flesh. Naganathan et al. (2008) explored the use of visible/near-infrared hyperspectral imaging based on the textural features for beef tenderness classification. In this study, PCA was implemented to determine the PC images that had eigenvalues significantly greater than zero. Then eight texture features (mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation) extracted from these selected images (first five PC images) were employed to classify beef tenderness into three categories, namely, tender, intermediate, and tough. In another study, Liu et al. (2010) developed Gabor filter-based hyperspectral imaging technique for categorization of pork quality and an average accuracy of 84±1% for pork quality level classification was obtained by using the textural features at 5 hybrid PC images. The use of PCA for texture feature selection was also realized in differentiating fresh and frozen-thawed fish fillets (Zhu et al., 2012). A total of 36 textural feature variables were extracted by GLCM from the first three PC images. As a result, the average correct classification rate reached 97.22% for predicting samples, which was based on the combination of spectral and textural variables. While the majority of studies were focused on the texture feature selection for classification, an attempt has been made to use texture

features for quality assessment of lamb meat. A number of 16 textural parameters at two selected images were used to investigate the relationship between drip loss and textural information (Kamruzzaman et al., 2012).

Although the stepwise regression algorithm has been widely employed in wavelength selection as discussed before. However, little information is available for its application in identifying important texture features. Choudhary et al. (2009) made an attempt to applying stepwise regression for selecting texture features from the hyperspectral images of wheat. For each slice of hypercube image, 20 energy and 20 entropy features were computed from the wavelet coefficients at each orientation within each level using the given equations. Thus the total number of texture features obtained from a hypercube of 75 slices was 3000. As a result of stepwise regression, only the top 100 texture features were used for classification (eight classes) of wheat and a satisfactory average classification accuracy of 99.1% was achieved by using statistical and back propagation neural network.

On the other hand β-regression coefficients method is also available for selecting textural features. For the prediction of beef tenderness, Naganathan et al. (2008) explored the use of texture features from near-infrared hyperspectral images. In this study, β regression coefficients of PLSR model was firstly implemented to reduce the number of images, leading to the selection of the images with high regression coefficient as the important images for further analysis of texture. An accuracy of 77% for classification (tenderness, intermediate and tough) was obtained

using 4.5% of original textural features of hyperspectral imaging, which proved that β-regression coefficients method was very effective for its application in texture feature selections.

3.3 Differences between Spectral Feature Selection Methodologies and Image Feature Selection Methodologies

The information of hyperspectral image forms a three-dimensional data cube, which contains one-dimensional spectral information (spectral feature) as well as two-dimensional spatial information (image feature) (Elmasry et al., 2012). Specially, the spectral features present the spectral values or variations at different wavelengths from an independent pixel (x_i, y_j) while the image features (such as texture features and shape features) reveal the variation and relationship among adjacent pixels from different directions at a certain wavelength. Therefore, two dimensions need to take into consideration while conducting image feature extraction or selection due to the various directions, which is more complex than that in spectral feature selection. In details, the relevance between two-dimensional features and prediction model should be analyzed before converting two-dimensional features into feature vectors by a variety of mapping methods, and then comes to the one-dimension feature selection (Guyon & Elisseeff, 2003).

The feature selection methodologies for hyperspectral image were divided into two parts in this paper, namely spectral feature selection methodologies and image feature selection

methodologies. On one hand, the two-dimensional signal (image feature) selection is essentially different from the one-dimensional signal (spectral feature) selection in space despite some similar principles of selection methods. On the other hand, more research efforts should be focused on image feature selection and more advanced feature selection methods should be applied to select image features for better usage of HSI.

4. Conclusions and Future Trends

In recent years, advanced feature selection algorithms have been popularly applied to extract the most important features from HSI for food quality evaluation and analysis. All these studies suggest that a robust prediction and computing efficiency can be obtained by using selected features (both for spectral and image features) rather than full features, which is an indication of the importance of feature selection in HSI. The fundamental principles and applications of several feature algorithms based on the searching strategy (complete search, heuristic search and random search) are described in details in this review. Due to the diversity of fundamental principles from each searching strategy method, applications and characteristics vary greatly from the complete search to random search. Currently, there is no consensus on the merits among the search strategy algorithms since different results may be acquired depending on different feature selection methods. Generally speaking, several elements should be considered while choosing feature selection methods such as the nature of the problem, the size of the dataset,

predicting accuracy and complexity. Moreover, the limitations of these searching strategies are also discussed in the current review. Complete search methods might not be suitable for analyzing HSI data because they are exhaustive and time-consuming. Heuristic search methods are the best developed and widely used feature selection methods in HSI due to its high efficiency and low complexity. However, for the case of the data with high complexity, the heuristic search is not able to produce a satisfactory result while the random search may be helpful to find an approximate optimal subset.

In order to further develop hyperspectral imaging techniques for food quality detection, in future more feature selection methods that are currently not used for selecting features from HSI data should be introduced to the food field. There is also a need to integrate existing methods by taking the advantages of different kinds of feature selection methods. Moreover, more advanced feature selection methods that with high efficiency and high accuracy should be proposed to extract the most important features from HSI.

Acknowledgements

The authors gratefully acknowledge the financial support from Guangdong Province Government (China) through the program of "Leading Talent of Guangdong Province (Da-Wen Sun)". This research was also supported by National Key Technologies R&D Program (2014BAD08B09). Specially thanks to Nannan Wang and Zhenjie Xiong from South China University of Technology for their kind suggestions.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433-459.

Araújo, M.U.C., Saldanha, T.C. B., Galvão R.K. H., Yoneyama, T., Chame, H. C. & Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems,* 57, 65-73.

Ariana, D. P., & Lu, R. (2010). Evaluation of internal defect and surface color of whole pickles using hyperspectral imaging. *Journal of Food Engineering, 96*(4), 583-590.

Babai, L. (1979). Monte-Carlo algorithms in graph isomorphism testing. *Université tde Montréal Technical Report,* D. M. S. No. 79-10.

Bajwa, S. G., Bajcsy, P., Groves, P., & Tian, L. F. (2004). Hyperspectral image data mining for band selection in agricultural applications. *Information & Electrical Technologies Division of ASAE, 47*(3), 895-907.

Barbin, D., Elmasry, G., Sun, D.-W., & Allen, P. (2012a). Near-infrared hyperspectral imaging for grading and classification of pork. *Meat Science, 90*(1), 259-268.

Barbin, D. F., ElMasry, G., Sun, D.-W., & Allen, P. (2012b). Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging. *Analytica Chimica Acta, 719*, 30-42.

Barbin, D. F., ElMasry, G., Sun, D.-W., Allen, P., & Morsy, N. (2013a). Non-destructive assessment of microbial contamination in porcine meat using NIR hyperspectral imaging. *Innovative Food Science & Emerging Technologies, 17*, 180-191.

Barbin, D. F., ElMasry, G., Sun, D.-W., & Allen, P. (2013b). Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging. *Food Chemistry, 138*(2-3), 1162-1171.

Burger, J., & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems, 108*(1), 13-22.

Centner, V., Massart, DL., Noord, O., Jong, S, Vandeginste, B. M., & Sterna, C. (1996). Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry, 68*, 3851-3858.

Chao, K., Mehl, P. M., & Chen, Y. R. (2002). use of hyper-multi-spectral imaging for detection of chicken skin tumors *Applied Engineering in Agriculture, 18*(1), 113-119.

Cheng, J., Sun, D.-W., Zeng, X.-A., & Liu, D. (2013). Recent advances in methods and techniques for freshness quality determination and evaluation of fish and fish fillets: A review. *Critical Reviews in Food Science and Nutrition*(accepted), DOI:10.1080/10408398.2013.769934.

Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems.*78 (1-2),

103-112.

Choudhary, R., Mahesh, S., Paliwal, J., & Jayas, D. S. (2009). Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosystems Engineering, 102*(2), 115-127.

Cluff, K., Naganathan, G. K., Subbiah J., Lu R. Calkins, C. R., & Samal, A. (2008). Optical scattering in beef steak to predict tenderness using hyperspectral imaging in the VIS-NIR region. *Sensing & Instrumentation for Food Quality and Safety.* 2:189-196.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1*, 131-156.

Di W. Hong X., Cai J.B., Huang, Z. H., & He, H. (2009). Classifying the species of exopalaemon by using visible and near infrared spectra with uninformative variable elimination and successive projections algorithm. *Journal of Infrared Millim. Waves, 28*(6), 423-427.

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 nutrition*. 52(8): 689-711.

ElMasry, G., Iqbal, A., Sun, D.-W., Allen, P., & Ward, P. (2011a). Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. *Journal of Food Engineering, 103*(3), 333-344.

ElMasry, G., Sun, D.-W., & Allen, P. (2011b). Non-destructive determination of water-holding

capacity in fresh beef by using NIR hyperspectral imaging. *Food Research International, 44*(9), 2624-2633.

ElMasry, G., Sun, D.-W., & Allen, P. (2012). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering, 110*(1), 127-140.

ElMasry, G., Sun, D.-W., & Allen, P. (2013). Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging. *Journal of Food Engineering, 117*(2), 235-246.

ElMasry, G., Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering, 81*(1), 98-107.

ElMasry, G., Wang, N., & Vigneault, C. (2009). Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks. *Postharvest Biology and Technology, 52*(1), 1-8.

Feng, Y.-Z., & Sun, D.-W. (2013a). Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of Pseudomonasloads in chicken fillets. *Talanta, 1*(57).

Feng, Y. Z., & Sun, D.-W., (2013b). Determination of total viable count (TVC) in chicken breast fillets by near-infrared hyperspectral imaging and spectroscopic transforms. *Talanta, 105*, 244-249.

Feng, Y. Z., ElMasry, G., Sun, D.-W., Scannell, A. G., Walsh, D., & Morcy, N. (2013). Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of Enterobacteriaceae on chicken fillets. *Food Chemistry, 138*(2-3), 1829-1836.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytical Chimica Acta, 185*(0), 1-17.

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 in Food Science & Technology, 18*(12), 590-598.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine learning Research, 3*(03), 1157-1182.

Haralick, R.M., Shanmugam,K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3, 610-621.

Jain, A., & Zongker, D., Feature selection: evaluation, application, and small sample performance (1997), *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2), 153-158.

Kamruzzaman, M., ElMasry, G., Sun, D.-W., & Allen, P. (2011). Application of NIR hyperspectral imaging for discrimination of lamb muscles. *Journal of Food Engineering, 104*(3), 332-340.

Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D.-W., & Allen, P. (2012a). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies, 16*, 316-325.

Kamruzzaman, M., ElMasry, G., Sun, D.-W., & Allen, P. (2012b). Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. *Innovative Food Science & Emerging Technologies, 16*, 218-226.

Kamruzzaman, M., ElMasry, G., Sun, D.-W., & Allen, P. (2012c). Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Analytical Chimica Acta, 714*, 57-67.

Kamruzzaman, M., ElMasry, G., Sun, D.-W., & Allen, P. (2013a). Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. *Food Chemistry, 141*(1), 389-396.

Kamruzzaman, M., Sun, D.-W., ElMasry, G., & Allen, P. (2013b). Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta, 103*, 130-136.

Karnopp, D. C. (1963). Random search techniques for optimization problems. *Automatica, 1(2-3)*, 111-121.

Konda Naganathan, G., Grimes, L. M., Subbiah, J., Calkins, C. R., Samal, A., & Meyer, G. E. (2008). Partial least squares analysis of near-infrared hyperspectral images for beef tenderness prediction. *Sensing and Instrumentation for Food Quality and Safety, 2*(3),

178-188.

Leardi, R., & Gonzalez, A. L. (1998). Genetic algorithms applied to feature selection in PLS regression:how and when to use them. *Chemometrics and Intelligent Laboratory Systems, 41* 195–207.

Li, H., Liang, Y., Xu, Q., & Cao, D. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytical Chimistry Acta, 648*(1), 77-84.

Li, J., Rao, X., & Ying, Y. (2011). Detection of common defects on oranges using hyperspectral reflectance imaging. *Computers and Electronics in Agriculture, 78*(1), 38-48.

Liu, H., Member, S., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering, 17*(4), 491-502.

Liu, L., Ngadi, M. O., Prasher, S. O., & Gariépy, C. (2010). Categorization of pork quality using Gabor filter-based hyperspectral imaging technology. *Journal of Food Engineering, 99*(3), 284-293.

Richard, E. K. (1990). Real-time heuristic search, *Artificial Intelligence*, 42(2-3), 189-211.

Mallat,S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence,11(7), 674–693.

Naganathan, G. K., Grimes, L. M., Subbiah, J., Calkins, C. R., Samal, A., & Meyer, G. E. (2008).

Visible/near-infrared hyperspectral imaging for beef tenderness prediction. *Computers and Electronics in Agriculture, 64*(2), 225-233.

Nakariyakul, S., & Casasent, D.(2004) Hyperspectral feature selection and fusion for detection of chicken skin tumors, *Proceeding of SPIE Volume 5271, Monitoring Food Safety, Agriculture, and Plant Health, 128*

Nakariyakul, S., & Casasent, D. P. (2007). Adaptive branch and bound algorithm for selecting optimal features. *Pattern Recognition Letters, 28*(12), 1415-1427.

Narendra, P., Fukunaga, K., 1977. A branch and bound algorithm for feature subset selection. IEEE Transaction on Computers. C-26(9), 917-922.

Qiao, J., Ngadi, M. O., Wang, N., Gariépy, C., & Prasher, S. O. (2007a). Pork quality and marbling level assessment using a hyperspectral imaging system. *Journal of Food Engineering, 83*(1), 10-16.

Qiao, J., Wang, N., Ngadi, M. O., Gunenc, A., Monroy, M., Gariepy, C., & Prasher, S. O. (2007b). Prediction of drip-loss, pH, and color for pork using a hyperspectral imaging technique. *Meat Science, 76*(1), 1-8.

Rajkumar, P., Wang, N., Eimasry, G., Raghavan, G. S. V., & Gariepy, Y. (2012). Studies on banana fruit quality and maturity stages using hyperspectral imaging. *Journal of Food Engineering, 108*(1), 194-200.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in

bioinformatics. *Bioinformatics, 23*(19), 2507-2517.

Sone, I., Olsen, R. L., Sivertsen, A. H., Eilertsen, G., & Heia, K. (2012). Classification of fresh Atlantic salmon (Salmo salar L.) fillets stored under different atmospheres by hyperspectral imaging. *Journal of Food Engineering, 109*(3), 482-489.

Talens, P., Mora, L., Morsy, N., Barbin, D. F., ElMasry, G., & Sun, D.-W. (2013). Prediction of water and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging. *Journal of Food Engineering, 117*(3), 272-280.

Tao, F., Peng, Y., Li, Y., Chao, K., & Dhakal, S. (2012). Simultaneous determination of tenderness and Escherichia coli contamination of pork using hyperspectral scattering technique. *Meat Science, 90*(3), 851-857.

Vemuri, V. (1988). Artificial neural networks. IEEE Computer Society Technology Series, 26.

Wang, S., Huang, M., & Zhu, Q. (2012). Model fusion for prediction of apple firmness using hyperspectral scattering image. *Computers and Electronics in Agriculture, 80*, 1-7.

Wang, W., Li, C., Tollner, E. W., Gitaitis, R. D., & Rains, G. C. (2012). Shortwave infrared hyperspectral imaging for detecting sour skin (Burkholderia cepacia)-infected onions. *Journal of Food Engineering, 109*(1), 38-48.

Wu, D., Shi, H., Wang, S., He, Y., Bao, Y., & Liu, K. (2012a). Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system. *Analytical Chimistry Acta, 726*, 57-66.

Wu, D., Sun, D.-W., & He, Y. (2012b). Application of long-wave near infrared hyperspectral imaging for measurement of color distribution in salmon fillet. *Innovative Food Science & Emerging Technologies, 16*, 361-372.

Wu, D., Wang, S., Wang, N., Nie, P., He, Y., Sun, D.-W., & Yao, J. (2012c). Application of time series hyperspectral imaging (TS-HSI) for determining water distribution within beef and spectral kinetic analysis during dehydration. *Food and Bioprocess Technology*.

Wu, D., & Sun, D.-W. (2013a). Application of visible and near infrared hyperspectral imaging for non-invasively measuring distribution of water-holding capacity in salmon flesh. *Talanta, 116*, 266-276.

Wu, D., & Sun, D.-W., (2013b). Potential of time series-hyperspectral imaging (TS-HSI) for non-invasive determination of microbial spoilage of salmon flesh. *Talanta, 111*, 39-46.

Wu, J., Peng, Y., Li, Y., Wang, W., Chen, J., & Dhakal, S. (2012). Prediction of beef quality attributes using VIS/NIR hyperspectral scattering imaging technique. *Journal of Food Engineering, 109*(2), 267-273.

Xing, J., Bravo, C., Jancsók, P. T., Ramon, H., & De Baerdemaeker, J. (2005). Detecting bruises on 'Golden Delicious' apples using hyperspectral imaging with multiple wavebands. *Biosystems Engineering, 90*(1), 27-36.

Xing, J., & De Baerdemaeker, J. (2005). Bruise detection on 'Jonagold' apples using hyperspectral imaging. *Postharvest Biology and Technology, 37*(2), 152-162.

Xu, L., & Zhang, W.-J. (2001). Comparison of different methods for variable selection. *Analytical Chimstry Acta, 446*, 477-483.

Zhang, X., Liu, F., He, Y., & Gong, X. (2013). Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. *Biosystems Engineering, 115*(1), 56-65.

Zhou, Z., Li, X., Gao, H., Tao, H., Li, P., & Wen, D. (2012). Comparision of different variable selection methods on potato dry matter detection by hyperspectral imaging technology. *Journal of Agricultural Machinery, 43*(2), 129-133.

Zhu, F., Zhang, D., He, Y., Liu, F., & Sun, D.-W. (2012). Application of visible and near infrared hyperspectral imaging to differentiate between fresh and frozen–thawed fish fillets. *Food and Bioprocess Technology*,6, 2931-2937.

Table 1. Applications of feature selection algorithms for analyzing spectral features of HSI data in foods.

Search Strategy	Algorithms	Applications	No. of PEW	References
Complete search	BB	Skin tumors of chicken	5	Nakariyakul & Casasent (2007)
Heuristic search	PCA	Bruise of apple	4	Xing et al. (2005)
	PCA	Defects of orange	6	Li et al. (2011)
	PCA	Defects of pickles	11	Ariana & Lu (2010)
	PCA	Sour skin of onions	4	Wang et al. (2012)
	PCA	Classification of pork quality	6	Qiao et al. (2007a)
	PCA	Muscle discrimination of lamb	6	Kamruzzaman et al. (2011)
	PCA	Discrimination of turkey hams	8	ElMasry et al. (2011a)
	Derivatives	TVC, PPC of pork	11, 10	Barbin et al. (2013a)
	Derivatives	Classification of pork quality	6	Barbin et al. (2012a)
	Derivatives	*Enterobacteriaceae* of chicken	4	Feng et al. (2013)
	Derivatives	Red meat classification	6	Kamruzzaman et al. (2012a)
	β-coefficients	Drip loss, pH, color of pork	6, 6, 6	Qiao et al. (2007b)
	β-coefficients	Protein, moisture, fat of pork	11, 7, 9	Barbin et al. (2013a)
	β-coefficients	WHC of beef	6	ElMasry et al. (2011b)
	β-coefficients	L^*, b^*, pH, tenderness of beef	6, 5, 24, 15	ElMasry et al. (2012)
	β-coefficients	Water, fat, protein of beef	8,7,10	ElMasry et al. (2013)
	β-coefficients	pH of lamb	8	Kamruzzaman et al. (2012c)
	β-coefficients	Water, fat, protein of lamb	6, 6, 6	Kamruzzaman et al. (2012b)
	β-coefficients	Adulteration of lamb	4	Kamruzzaman et al. (2013a)
	β-coefficients	*Enterobacteriaceae* of chicken	3	Feng et al. (2013)

	β-coefficients	Protein, water of hams	10, 6	Talens et al. (2013)
	β-coefficients	Classification of salmon	5	Sone et al. (2012)
	β-coefficients	pH of strawberry	8	ElMasry et al. (2007)
	β-coefficients	Maturity stages of banana	8	Rajkumar et al. (2012)
	β-coefficients	N, P, K of oilseed	12, 8, 5	Zhang et al. (2013)
	Stepwise	Tenderness, L*, a*, b* of beef	8, 7, 7, 8	Wu et al. (2012c)
	Stepwise	Tenderness of beef	7	Cluff et al. (2008)
	Stepwise	TVC of chicken	5	Feng & Sun (2013b)
	Stepwise	Tenderness, E. *coli* of pork	6, 5	Tao et al. (2012)
	UVE	Classification of prawns	12	Di et al. (2009)
	UVE	Firmness of apple	34	Wang et al. (2012)
	SPA	Water content of beef	6	Wu et al. (2012c)
	SPA	Tenderness of lamb	11	Kamruzzaman et al. (2013b)
	SPA	Moisture content of prawns	12	Wu et al. (2012a)
	SPA	L*, a*, b* of salmon	4, 6, 10	Wu et al. (2012b)
	SPA	Dry matters of potato	27	Zhou et al. (2012)
	ANN	Injuries of apple	5	ElMasry et al. (2009)
Random search	GA	Pseudomonas of chicken	5	Feng & Sun (2013b)
	CARS	TVC of salmon	8	Wu & Sun (2013a)
	CARS	WHC of salmon	12	Wu & Sun (2013b)

PEW: effective wavelengths for prediction.

BB: branch and bound; PCA: principal component analysis, Derivatives: spectral derivatives, β-coefficients: regression coefficients, Stepwise: stepwise regression, UVE: uninformative variable elimination, SPA: successive projections algorithm, ANN: artificial neural network; GA: genetic algorithms, CARS: competitive adaptive reweighted sampling.

Table 2. Applications of feature selection algorithms for analyzing image features of HSI data in foods.

Algorithms	Applications	No. of Image Features	References
PCA	Classification of beef	40	Naganathan et al. (2008)
	Categorization of pork	125	Liu et al. (2010)
	Classification of fish fillets	36	Zhu et al. (2012)
	Quality of lamb	32	Kamruzzaman et al. (2012)
Stepwise	Classification of wheat	100	Choudhary et al. (2009)
β-coefficients	Tenderness of beef	48	Naganathan et al. (2008)

PCA: principal component analysis, Stepwise: stepwise regression, β-coefficients: regression coefficients.

Table 3. A comparison of feature selection algorithms

Search strategy	Advantages	Disadvantages
Complete search	Guaranteeing optimal	Huge searching space
		Time-consuming
Heuristic search	Including various kinds of algorithms	Losing optimal subset
	Small searching space	Unsuitable for high complex problems
	More efficiency and low complexity	
Random search	Approximate optimal	Time-consuming
	High capability for complex problems	Unstable results

Figure Captions

Fig.1. The main procedure of feature selection.

Fig. 2. The solution tree for the basic BB algorithm.

Fig. 3. Example of SPA with J=6, M_{cal}=3 and k(0)=3.

Fig. 4. Flow chart of CARS algorithm.

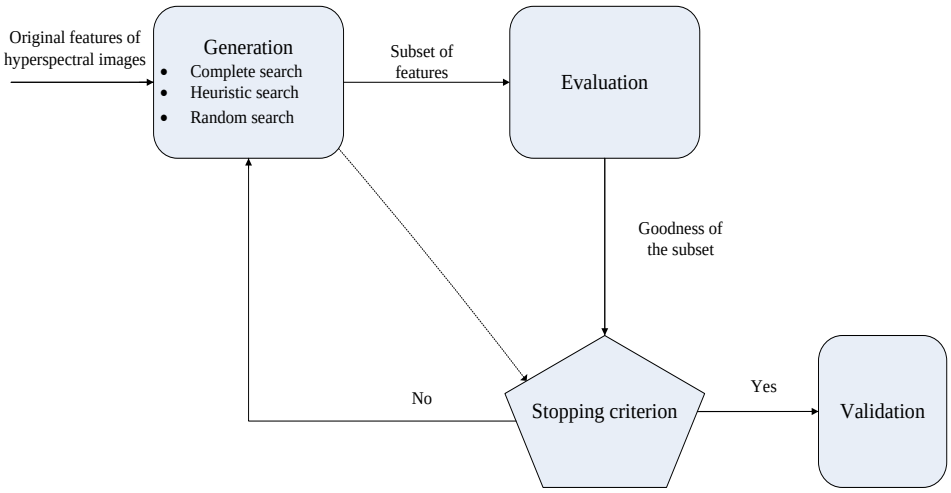

Fig. 1. The main procedure of feature selection.

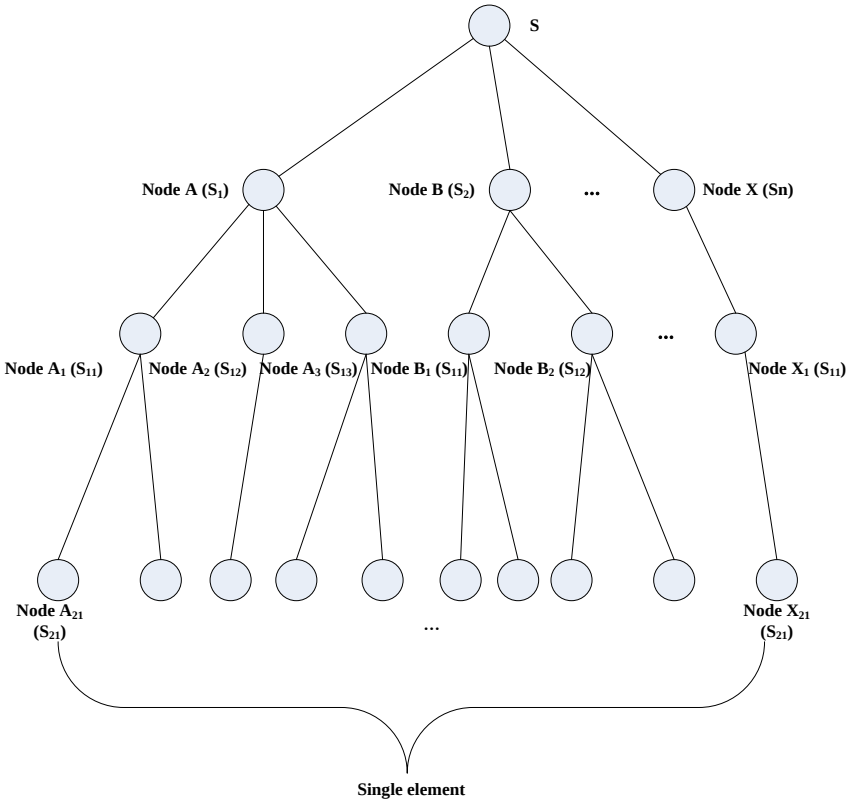

Fig. 2. The solution tree for the basic BB algorithm.

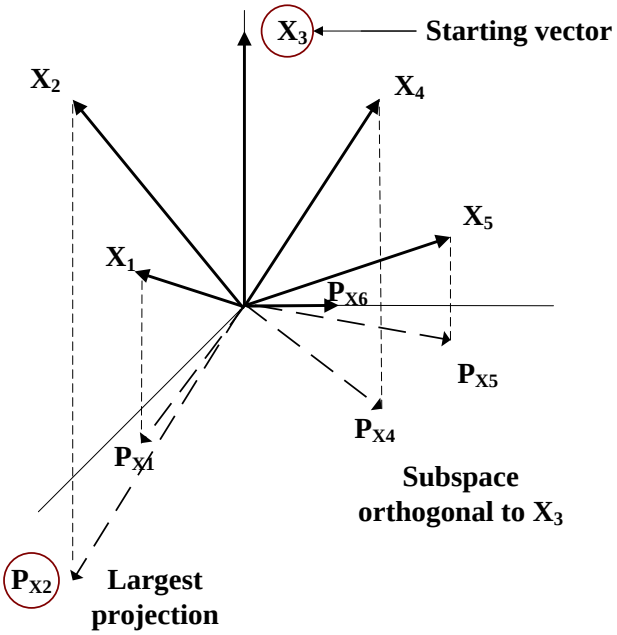

Fig. 3. Example of SPA with J=6, M_{cal}=3 and k (0) =3.

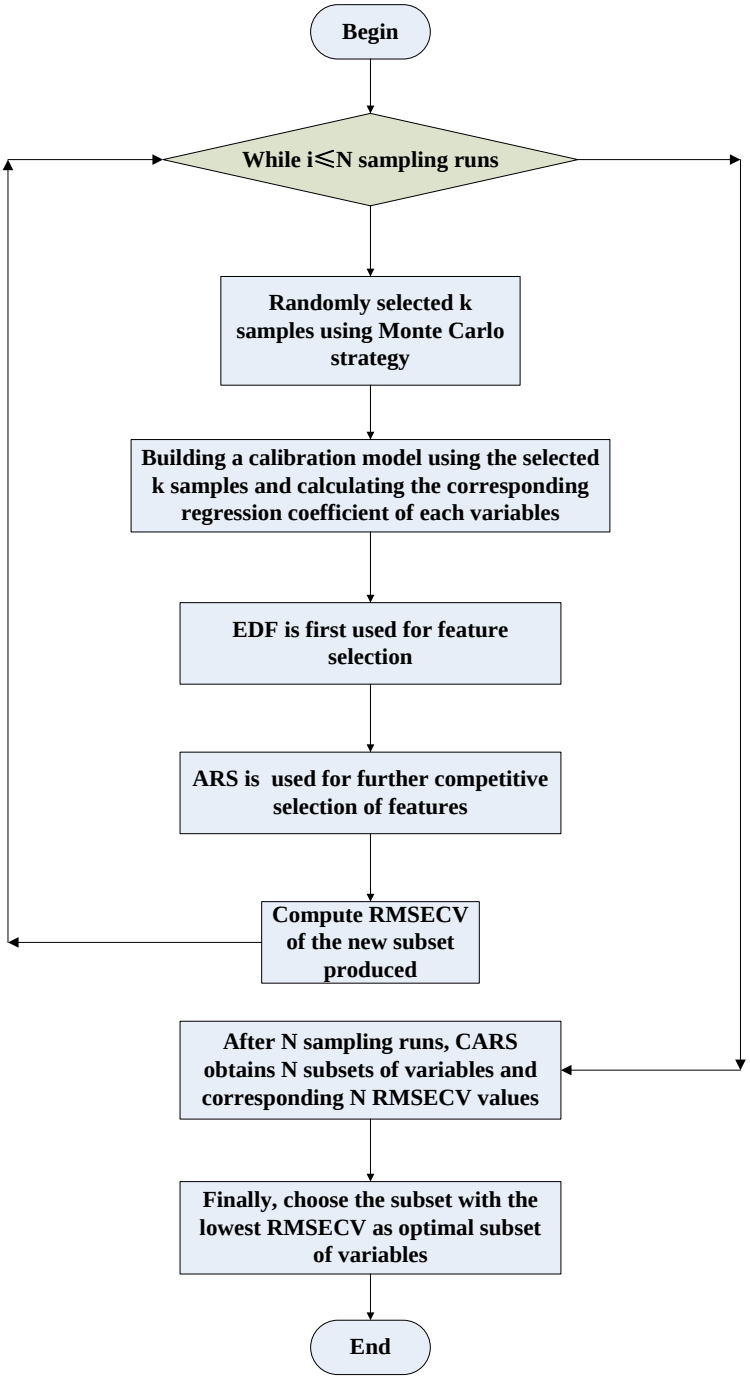

Fig. 4. Flow chart of CARS algorithm.