JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Classification of Modified Starches by Fourier Transform Infrared Spectroscopy Using Support Vector Machines

J. A. FERNÁNDEZ PIERNA,[†] P. VOLERY,[‡] R. BESSON,[‡] V. BAETEN,[§] AND P. DARDENNE*,[§]

Scientific Collaborator F.N.R.S., Statistics and Computer Sciences Department, University of Agronomical Sciences, Avenue de la Faculté 8, 5030 Gembloux, Belgium, Quality and Safety Department, Nestlé Research Centre, Vers-chez-les-Blanc, P.O. Box 44 1026 Lausanne, Switzerland, and Quality of Agricultural Product Department, Walloon Agricultural Research Centre (CRA-W), Chaussée de Namur no. 24, 5030 Gembloux, Belgium

The aim of this study was to compare the performance of different supervised discrimination methods based on IR data for the classification of starches according to the type of chemical modification undergone. The goal of the supervised classification methods is to develop classification rules. Representative samples of each group (known beforehand) were available, from which the relevant characteristics (chemical modification) were known. On the basis of a training data set, classification rules are determined, which can then be applied to classify new (unknown) samples.

KEYWORDS: Starches; FTIR; classification methods; chemometrics; SVM

INTRODUCTION

Starch is naturally found in all sorts of plants, located in their roots, stems, seeds, or fruits. Starch is a source of energy for the plants during dormancy and germination, and it serves similar purposes for human and animal nutrition.

The most common sources of food starch are corn, potato, wheat, tapioca, and rice. All starches are made up of amylose and amylopectin, with a ration that varies according to the starch source. Corn contains about 25−28% amylose, while high amylose corn can be as high as 80%. Tapioca has about 17% amylose, and waxy maize has virtually none.

The building blocks of starch are α-D and β-D glucose, which are linked together through enzymatic condensation occurring mainly between carbons 1 and 4, but occasionally between 1 and 6. Amylose refers to a linear chained homopolymer containing only α-1,4 linkages. The length of this chain will vary with plant source, but the average length will be between 500 and 2000 glucose units. Amylopectin refers to a branched polymer, due to occasional linkage between glucose units at carbons 1 and 6. This results in the development of a more massive molecule but with linear chain lengths of only 25−30 glucose units.

In the food industry, starch is used to modulate product characteristics such as texture, appearance, or stability, in a wide range of applications from baby foods to ice cream. In the unmodified or native form, starches have limited use in the food industry; therefore, starches undergo a process of modification (physical or chemical) to modulate their properties to provide the expected thickening, water binding, stabilizing, gelling effect or to improve mouth feel and shininess of the product. Starches can also be cross-linked in order to provide heat, acid, and shear tolerance.

Identification of raw materials is a requirement of the Good Manufacturing Practices, with the aim of ensuring product safety, raw material traceability, and consistent quality. Starches can be identified by means of traditional wet chemical methods, as described in the Compendium of Food Additive Specifications, Addendum 5 "Modified Starches" (*1*). These methods are time- and money-consuming, and they require skilled operators. They are not adequate for the rapid identification check to be performed in the food industry at the reception of raw materials or just before their use in production.

Various alternative methods such as near infrared spectroscopy (NIR), Fourier transform infrared spectroscopy (FTIR), nuclear magnetic resonance, X-ray fluorescence, or X-ray diffraction spectroscopy have already been successfully evaluated for the identification or the characterization of selected raw materials. Applying new powerful chemometric tools, FTIR spectroscopy has proved to be a promising technology for the identification of modified starches (*2*, *3*).

This study aimed at evaluating the potential of FTIR spectroscopy to identify modified starches in a food industry environment. For this purpose, different discrimination techniques were applied as follows: classical methods such as linear discriminant analysis (LDA) (*4*, *5*), quadratic discriminant analysis (QDA) (*5*), *k*-nearest neighbors (kNN) (*6*), soft independent modeling of class analogies (SIMCA) (*7*), partial least squares discriminant analysis (PLS-DA) (*8*), or artificial neural networks (ANN) (*9*, *10*) and more recent techniques such as support vector machines (SVM) (*11*−*16*).

* To whom correspondence should be addressed.
[†] University of Agronomical Sciences.
[‡] Nestlé Research Centre.
[§] Walloon Agricultural Research Centre (CRA-W).

10.1021/jf0501544 CCC: $30.25
Published on Web 00/00/0000

THEORY

For the application of all of the different classification methods, the data set was summarized by an *nxm* matrix $\boldsymbol{X}$ where *n* denotes the number of samples (starches) and *m* denotes the number of wavelengths. In such a way, when the samples belong to known classes, the data for each observation *i* consisted of a predictor vector $x_i = (x_{i,1},...x_{i,m})$ and a class level (response or dummy variable) $\boldsymbol{y}_i$. For K classes, the class labels y_i were defined to be integers ranging from 1 to *k*. Predictors were built from observations, which were known to belong to certain classes (i.e., from past experience). Such observations constituted the training set. The results (models) obtained using this training set may then be applied to a test set. Here, a full analysis was carried out including diagnostics, data reduction/selection, modeling, and validation of models.

Data Reduction. Because in IR data sets the number of variables is larger than the number of objects, most of the discrimination methods cannot be directly applied. Two methods for data reduction are applied here: the principal component analysis (PCA) (*17*) and the Fisher criterion (FC) (*18*).

PCA is used to reduce the data dimensionality by creating new orthogonal variables (scores or latent variables) that are linear combinations of the original measured ***x***-variables (absorbances at different wavelengths). Only the first new variables or PCs, accounting for most of the variance of the original data, contain meaningful information, while the last ones, which account for a small amount of variance, contain more noise and can be ignored.

PCA is also performed on the data in order to make a first homogeneity study. The scores ($\boldsymbol{T}$) can highlight clustering, trends, and outliers in the sample space in the data matrix. The loadings ($\boldsymbol{P}$) focus on the variation in the variable space, describing the influence of the variables on the scores $\boldsymbol{T}$. PCA is only able to identify large variability and is not always able to distinguish "between groups" and "within groups" variability. For that, the FC can be used. The FC describes the ratio of "between class" variance to "within class" variance that is helpful to decide which original variables have an important discriminating power.

For each variable *j*, the FC maximizes the distance between the means of the classes while minimizing the variance within each class. This criterion is described as:

$$FC_j = H_j/E_j \tag{1}$$

where $H_j = \sum_{i=1}^{K} n_i(\bar{x}_{ij} - \bar{x}_{.j})^2$ is the between class variance and $E = \sum_{i=1}^{K} (n_k - 1)s_{ij}^2$ is the within class variance, where n_i is the number of objects in class *i*, $\bar{x}_{ij}$ is the mean absorbance of the objects belonging to class *i* at the *j*-th wavelength, $\bar{x}_{.j}$ is the mean absorbance of the objects belonging to all classes at the *j*-th wavelength, and s_{ij} is the standard deviation of the absorbance of the objects belonging to class *i* at the *j*-th wavelength.

Discrimination. First, a distinction between supervised and unsupervised learning methods has to be done (*5*). A learning method is considered supervised if information about the structure/grouping of the objects is assumed to be known or at least partially known. This prior knowledge is then used in the analysis process. However, prior information about data is often not available, in which case the information needs to be inferred directly from the given spectral data in an unsupervised manner. An unsupervised learning method learns in the absence of a teacher signal that provides a prior knowledge of the true classes. Whereas unsupervised methods determine how a set of data clusters into functional groups, supervised methods determine what expression characteristics of a given data make it part of a given functional group. Supervised methods can use complex models that exploit the specific characteristics of the given functional group. The supervised classification methods can be divided into two groups: simply discrimination methods and class-modeling techniques. The aim of the simply discrimination methods is to find classification rules, which define optimal boundaries between all given classes by maximizing the difference between them. A new object will be assigned to one class, namely, to the class to which it is the most similar. The simplest method is the one based on the FC, the Fisher linear discriminant method (*19*). This method measures the weighted (by the intraclass scattering) separation between the classes. The larger the FC is, the larger the weighted distance between the classes. Methods such as LDA (*4*, *5*), QDA (*5*), or kNN (*6*) can also be included in this group. Also, classical techniques such as PLS-DA (*8*) or ANN (*9*, *10*) and a more recent approach, namely, SVM (*11*−*16*), are included in this simple discrimination group. A shortcoming of such methods is that the detection of a sample not belonging to any group becomes impossible. The class-modeling techniques can deal with this problem because they establish an individual model for each class based on its training set. For this kind of method, more emphasis is put on similarities between the objects within a class. An unknown object will only be assigned to a class, if it falls within the defined volume of that class. If it falls outside, it is considered as an outlier for that class. Methods such as SIMCA (*7*) can be included in this group.

All of these techniques (LDA, QDA, kNN, PLS, ANN, and SIMCA) have been widely described in the literature (*5*) showing their drawbacks and advantages. Here, the most recent technique, SVM, is described.

SVM. SVM is a supervised method that has been applied in the literature (*11*−*16*) to a large range of pattern recognition problems such as object recognition, face detection, or text categorization. The aim of SVM is to find an optimal hyperplane (classifier) that correctly separates objects of the different classes as much as possible. This is done by leaving the largest possible fraction of points of the same class on the same side and maximizing the distance of either class from the hyperplane. The classifier would be useful in recognizing new members of the class.

For a two class problem, SVM chooses a specific hyperplane among the many that can separate the data in order to avoid overfitting. This hyperplane is chosen as the one that maximizes the minimum distance from the hyperplane to the closest training points in both classes (margin). These points determining the hyperplane are called support vectors. The hyperplane is found by defining a mapping $[z = \phi(x_i)]$ that transforms the *d* dimensional input data x_i into a higher d' dimensional space (feature space).

A SVM model consists of finding an optimal solution for the quadratic programming problem:

$$\min\left(\frac{1}{2}||\alpha_i y_i||^2 + C\sum_{i=1}^{n}\xi_i\right) \tag{2}$$

under the constraints:

$$y_i\,[\alpha_i y_i \phi(x_i)^T + b] \geq 1 - \xi_i$$

$$\xi_i \geq 0, \text{ for } i = 1,...,n$$

where *n* is the number of training data points, y_i is the reference

Table 1. Description of the Sample Sets: Origin and Distribution in the Classes

	no. of samples				total	
name	United States	EU	total	removed samples total	training	test
native	11	27	38		28	10
E1401	4	0	4	4		
E1404	0	1	1	1		
E1411	0	1	1	1		
E1412	12	13	25		18	7
E1414	0	1	1	1		
E1422	22	35	57		42	15
E1440	5	1	6	6		
E1442	100	12	112		84	28
unknown	9	39	48	48		
E1450	0	1	1	1		
E1402	0	2	2	2		
E1420	0	1	1	1		
E1455	0	1	1	1		
total	163	135	298	66	172	60

or class value (± 1 for the training points x_i), α_i are the parameters learned from the data (Lagrange multipliers), b represents the offset parameter, C is the penalty to pay for accepting classification errors, and ξ_i represents the variable that measures this misclassification error. The first part of this equation explains for the maximization of the margin, and the second part explains for the minimization of misclassification error. Because solving this equation is difficult, due to the high dimensionality of the feature space, a kernel function is used to calculate the separating hyperplane. This function implicitly represents the construction of an optimal hyperplane in a high dimensional space and then returns to the original space as a nonlinear decision frontier. With an appropriately chosen kernel function, any consistent training set can be made separable. The kernel used here is a Gaussian radial basis function.

$$K(x_i,x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) \text{ for all pairs of objects} \quad (3)$$

where σ is the width of the Gaussian function and it reflects the degree of generalization.

The penalty C has to be added in order to take into account those samples that cannot be separated. When C is large, the second term of the eq 2 dominates, forcing the SVM toward a solution with the least training error and decreasing the amount of generalization. When C is decreased, more emphasis is placed on maximizing the margin and emphasizing the generalization. This generalization could also be obtained by increasing the value of σ in the Gaussian function.

The points that determine the hyperplane are the support vectors, and they correspond to the points for which α_i is positive. The other points are not used and $\alpha_i = 0$. Different algorithms have been proposed in the literature to perform SVM for classification. The Lin's Lib SVM v2.33 algorithm (*20*) is used in this study.

MATERIALS AND METHODS

Samples. Two hundred ninety-eight starch samples were collected in various factories located in the United States and in Europe. The distribution of their true classification and of their origin is presented in **Table 1**.

Sixty-six samples were removed from the data set, because insufficient information on their classification was available or because of the too low number of samples in some classes. Therefore, the final data set contained 232 starch samples of four different classes: the class "Nativ" consisted of 38 samples, the class "E1412" consisted of 25 samples, the class "E1422" consisted of 57 samples, and the class "E1422" consisted of 112 samples.

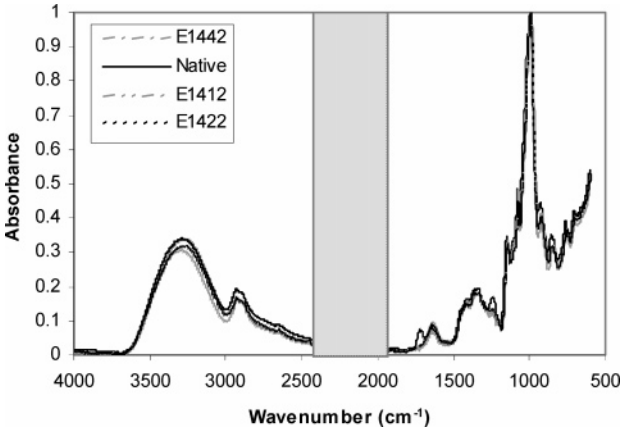

Figure 1. Typical FTIR spectra of starches modified or not recorded using an ATR accessory.

This data set was split into two subsets: a training set containing 75% of the objects from each class. This subset was later used for the construction of the discrimination models, and a test set, containing 25% of the objects from each class, was used for validation. Thus, 172 samples were available for the model development, and 60 samples were used for validation. Moreover, an independent data set was available, containing 36 samples consisting of 32 samples coming from the EU and four from the United States.

Spectra Acquisition. Midinfrared spectra (4000 to 600 cm^{-1} at 1 cm^{-1} data intervals) were collected with a spectral resolution of 4 cm^{-1} on a Perkin-Elmer Spectrum 2000 FTIR spectrometer (Perkin-Elmer Corp., Norwalk, CT) equipped with an ATR system Specac MKII GoldenGate (Specac Inc., Smyrna, GA) positioned so that the incident angle was 45°. The spectrophotometer was fitted with a wire coil operated at 1350 K as a IR light source, a potassium bromide beam splitter, and a DTGS detector.

Sample storage, sample preparation, and data acquisition were carried out at 25 °C. Data acquisition was performed over several days, and samples were taken in random order. After thorough mixing, a portion of the powder sample was positioned on the ATR diamond surface. A pressure of 11000 psi was applied on the sample by means of a pressure clamp. Each spectrum represents the average of 16 scans radioed against the background, which was collected with the empty ATR accessory under the same conditions at the beginning of each day of analysis. Between the samples, the ATR surface was thoroughly cleaned with water and alcohol. As no thermal equilibration was required, the whole spectra acquisition procedure took 3 min per sample.

Data acquisition was carried out using Spectrum Quant+ version 4.51.02 (Perkin-Elmer Corp.). The reflectance spectra were transformed into absorbance prior to their transfer to Matlab 7.0.

Typical spectra of starches (not modified and various types of modification) are presented in **Figure 1**. Some differences can be observed between the spectra. However, they are very small; therefore, the use of chemometric tools is necessary to build a powerful, robust, and discriminating model.

Prior to data analysis and model development, all spectra were trimmed. The range between 2402 and 1901 cm^{-1} was removed from the spectra, because it did not contain any useful information and was rather noisy (gray area in **Figure 1**). The resulting spectra contained 1901 data points: from 4000 to 2401 cm^{-1} and from 1900 to 600 cm^{-1}.

Software. All computations, chemometric analyses, and graphics were carried out with programs developed in Matlab v7.0 (The Mathworks, Inc., Natick, MA).

Model Implementation. In this study, the performance of different classification methods was compared. Within the training set, leave-one-out cross-validation was carried out to optimize the model, i.e., to find the number of variables, with which one obtains the best

Table 2. Performance of the Different Classifiers

			success rate	
method	preprocessing	no. parameters	training	test
LDA	raw	16[a]	0.9012	0.8202
	snv	12	0.8663	0.8119
	first deriv	11	0.8779	0.7548
QDA	raw	5[a]	0.8547	0.7351
	snv	4	0.8430	0.6923
	first deriv	10	0.8837	0.7298
1NN	raw	2[a]	0.7791	0.6000
	snv	7	0.8081	0.6262
	first deriv	4	0.8256	0.6797
3NN	raw	3[a]	0.8023	0.6000
	snv	4	0.8430	0.5542
	first deriv	4	0.8372	0.6797
SIMCA	raw	3/3/3/3[b]	0.8006	0.7423
	snv	2/2/2/2	0.7560	0.7137
	first deriv	1/1/1/1	0.7054	0.7631
SVM	raw	100/1[c]	0.9826	0.8333
	snv	100/0.035	0.9767	0.8500
	first deriv	10000/1	0.9012	0.8333
ANN	raw	15/3, 14/1, 11/1, 5/1[d]	0.8994	0.7653
	snv	16/2, 9/4, 11/1, 6/4	0.9336	0.8284
	first deriv	5/4, 13/2, 4/3, 7/5	0.8530	0.7971
PLS-DA	raw	15[e]	0.9445	0.7649
	snv	15	0.9563	0.8006
	first deriv	15	0.9673	0.7726

[a] Number of PCs. [b] Number of PCs for each of the individual PC models. [c] C/σ. [d] For each model: number of inputs (PCs)/number of hidden nodes. [e] Number of latent variables.

classification rate. To do that, the reduction of the feature space was performed using a combination of PCA and the FC, i.e., the PCs were ranked according to the FC in order to decide which PCs have an important discriminating power.

For PLS-DA, the number of latent variables was selected as the one that drove to the minimum RMSE (root-mean-square error). In ANN, the data set was divided into three subsets: a training set, a monitoring set, and a test set. The model was built with the training set, and the monitoring set was necessary to avoid overtraining. For SVM, the C and σ parameters were optimized using the optimization separate training and validation subsets. The optimal parameter settings C and σ were selected as the values that give the minimum RMSE and the maximum classification rate.

The classification or success rate is defined as:

$$\text{success rate} = \frac{\left(\sum_{i=1}^{K} \frac{\text{correctly classified samples in class } i}{\text{total number of samples in class } i}\right)}{K} \quad (4)$$

with K being the number of classes. A success rate of, for instance, 0.9012 indicates that 90.12% of the objects are correctly classified. This study was carried out on the raw data and on the data using two different preprocessing techniques: the SNV (standard normal variate) transformation and the first derivative.

RESULTS AND DISCUSSION

From **Table 2** and **Figure 2**, the performance of the different classifiers can be compared. In this table, the success rates obtained for the training set with cross-validation and the success rates for the classification of the test set are summarized.

From these results, several conclusions can be drawn. Among the simple discrimination methods, LDA performed better than QDA. This is an indication that the assumptions for LDA (normally distributed data, equal variance−covariance matrix for all classes) were sufficiently well-fulfilled. The performance of QDA was somewhat worse because more parameters had to be estimated; therefore, more objects per class were needed. LDA was also applied on the raw data without using PCA and by selecting the variables by the FC. Working in such a way, the results were not as good as the ones using the principal components (PCs) with a success rate for the training and for the test set of 0.794 and 0.755, respectively.

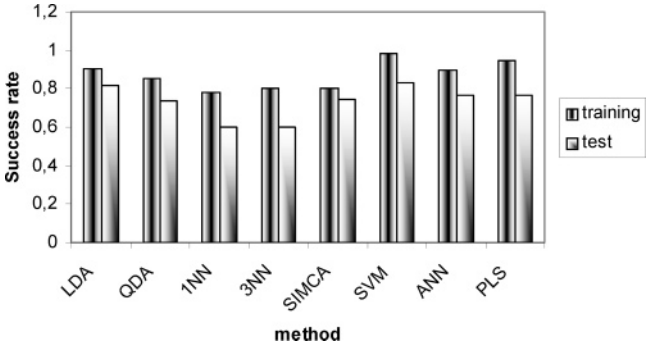

Figure 2. Performance of the different classifiers for the raw data.

kNN is a method that does not consider the shape of classes and defines similarities to a class only according to similarities of individual points. The chance of classifying an object wrongly is higher with 1NN than with 3NN. In the case of 3NN, more neighbors were included and the results became more reliable. In the case of 3NN, the objects were classified according to the majority rule. If an object is classified an equal number of times in several classes, then it will not be classified at all, and this explains some of the failures obtained with 3NN.

In SIMCA, a classification model (PC model) was constructed for each class individually. Here, it was arbitrarily decided to retain only the PCs containing more than 1% of the total variance for modeling with a maximum of three PCs for each model. For every model, a success rate was calculated. An important factor for the classification using SIMCA is the number of PCs to be included in the different models.

The best results were obtained with PLS-DA, ANN, and mainly with SVM. In all cases, the correct selection of the tuning parameters (number of latent variables for PLS-DA, number of input factors, and hidden nodes for ANN and C and σ for SVM) was critical to get good predictive performance. When looking at the success rates obtained with the different discrimination equations for the test set, one can conclude that several objects were misclassified. Different reasons were possible for that. The first one could be the natural heterogeneity and the dimension of the data set. When using leave-one-out cross-validation, one object was left out at the time, in such a way that if an extreme sample was left out, the remaining objects could not span the same space anymore, and as result, the object left out would not be classified in this class.

Another reason could be the selection of the optimal parameters for all of the methods (number of PCs, latent variables, C and σ, input, and hidden nodes). This is a critical point for all of the discrimination techniques. The classification results showed, for instance, that models established with different number of parameters lead to different success rates (not shown).

Preprocessing of the spectra did not influence much the results here, but it seemed nevertheless to be useful. SNV and first derivative decrease physical spectral information due to particle size, so that the models were mainly built based on chemical spectral information. Moreover, it increased the between class variance.

Once the different classification models were constructed and validated, they could be used for the prediction of the new data set proposed here. The results for this data set using all SVM

Table 3. Performance of the SVM Method on a New Data Set[a]

sample	name	real	predicted SVM
1	N003	Nativ	Nativ
2	N004	Nativ	Nativ
3	N009	Nativ	Nativ
4	N010	Nativ	Nativ
5	N012	Nativ	Nativ
6	N037	E1412	Nativ
7	N045	E 1442	E1442
8	N047	Nativ	Nativ
9	N071	E 1442	E1442
10	N078	Nativ	Nativ
11	N079	Nativ	E1412
12	N080	Nativ	Nativ
13	N081	E 1442	E1442
14	N082	Nativ	Nativ
15	N083	Nativ	E1442
16	N084	E 1442	E1442
17	N085	Nativ	Nativ
18	N089	Nativ	Nativ
19	N091	Nativ	Nativ
20	N093	Nativ	Nativ
21	N099	Nativ	Nativ
22	N100	E 1412	Nativ
23	N103	Nativ	E1412
24	N110	Nativ	E1412
25	N116	E 1442	E1442
26	N120	E 1442	E1442
27	N121	E 1442	E1442
28	N122	E 1442	E1442
29	N123	E 1422	E1422
30	N124	Nativ	Nativ
31	N128	Nativ	Nativ
32	N129	Nativ	Nativ
33	J7139-002	E1442	E1442
34	J7139-005	E1442	E1412
35	J7139-006	E1442	E1442
36	J7139-019	E1442	E1412

[a] The first 32 samples correspond to EU data, and the last four samples correspond to U.S. data.

are shown in **Table 3** (only raw results are shown). The first 32 samples correspond to the samples coming from the EU and the four last samples are the samples coming from the United States. The first and second columns represent the sample number and the sample name, respectively, the third column shows, for each object, the real group where the sample belongs, and the last column represents the group attribution for each sample using SVM. For this data set, eight samples were misclassified that correspond to 22% of the data, and a 78% of correct classification was obtained. These results correspond in a good way to the results obtained with the validation data set (see **Table 2**). If the results were studied according to the origin of the samples, then 81.25 and 50% of the EU samples and U.S. samples, respectively, were correctly classified. The low percentage obtained for the U.S. data set could be due to the low number of samples available for prediction (only four samples) and that the U.S. data were collected differently as the EU samples.

In conclusion, the results of this study showed that the different discrimination methods based on IR data can be effective tools for the classification of starches according to the type of chemical modification undergone. All of the classical discrimination methods used in this study could be successfully applied to classification problems, but the best results were the ones obtained using the more recent SVM technique. SVM is very robust against high dimensionalities and shows a very good performance in all of the cases. These properties, combined with a clear theoretical background, make SVM a good candidate to be applied to any complex industrial pattern recognition problem.

ACKNOWLEDGMENT

This work was performed within the context of a F.N.R.S. collaboration with the Statistics and Computer Sciences Department of the University of Agronomical Sciences in Gembloux, Belgium. We thank Prof. Claustriaux and Prof. Palm from that department for their valuable advice and support.

LITERATURE CITED

(1) Compendium of Food Additive Specifications. Addendum 5 (FAO Food and Nutrition Paper—52 Add. 5), Modified Starches, Joint FAO/WHO Expert Committee on Food Additives, AFO, Rome, 1997. http://www.fao.org/docrep/W6355E/w6355e0o.htm.

(2) Dolmatova, L.; Ruckebusch, C.; Dupuy, N.; Huvenne, J.-P.; Legrand, P. Identification of modified starches using infrared spectroscopy and artificial neural network processing. *Appl. Spectrosc.* **1998**, *52* (3), 329−338.

(3) Dupuy, N.; Laureyns, J. Recognition of starches by Raman spectroscopy. *Carbohydr. Polym.* **2002**, *49*, 83−90.

(4) Balakrishnama, S.; Ganapathiraju, A. Linear discriminant analysis—A brief tutorial. Institute for Signal and information processing, March 2, 1998. http://www.isip.msstate.edu/publications/reports/isip_internal/1998/linear_discrim_analysis/.

(5) Wu, W.; Mallet, Y.; Walczak, B.; Penninckx, W.; Massart, D. L.; Heuerding, S.; Erni, F. Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Anal. Chim. Acta* **1996**, *329*, 257−265.

(6) Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: A Textbook*; Elsevier Science Publishers B. V.: Amsterdam, The Netherlands, 1988; Chapter 23, pp 395−397.

(7) Wold, S.; Sjostrom, M. *Chemometrics: Theory and Application*; Kowalski, B. R., Ed.; American Chemical Society: Washington, 1977; p 243.

(8) Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166−173.

(9) Despagne, F.; Massart, D. L. Tutorial review: Neural networks in multivariate calibration. *Analyst* **1998**, *123*, 157R−178R.

(10) Bishop, C. M. *Neural Networks for Pattern Recognition*; Clarendon Press: Oxford, 1995.

(11) Vapnik, V. N. *The Nature of Statistical Learning Theory*, 2nd ed.; Springer: New York, 2000.

(12) Belousov, A. I.; Verzakov, S. A.; von Frese, J. Applicational aspects of support vector machines. *J. Chemom.* **2002**, *16*, 482−489.

(13) The Kernel Machines Website. http://www.kernel-machines.org.

(14) Suykens, J. A. K.; van Gestel, T.; De Brabanter, J.; De Moor, B.; Vandewalle, J. *Least Squares Support Vector Machines*; World Scientific: Singapore, 2002.

(15) Fernández Pierna, J. A.; Baeten, V.; Michotte Renier, A.; Cogdill, R. P.; Dardenne, P. Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *J. Chemom.* **2004**, *18*, 1−9.

(16) Cogdill, R. P.; Dardenne, P. Least-squares support vector machines for chemometrics. *J. Near Infrared Spectrosc.* **2004**, *12* (2), 93−100.

(17) Martens, H.; Naes, T. *Multivariate Calibration*, 2nd ed.; Wiley: Chichester, United Kingdom, 1989; Vol. 1.

(18) Duda, R. O.; Hart, P. E. *Pattern Classification and Scene Analysis*; John Wiley & Sons Inc.: New York, 1973.

(19) Szeto, L. K.; Liew, A. W. C.; Yan, H.; Tang, S. S. *Gene Expression Data Clustering and Visualization Based on a Binary Hierarchical Clustering Framework*; Chen, Y.-P. P., Ed.; First Asia-Pacific Bioinformatics Conference, Adelaide, Australia, Conferences in Research and Practice in Information Technology, 2003; Vol. 19.

(20) Website. http://www.csie.ntu.edu.tw/∼cjlin/libsvm.

Received for review January 21, 2005. Revised manuscript received June 17, 2005. Accepted July 5, 2005.

JF0501544