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

ARTICLE *in* JOURNAL OF CHEMOMETRICS · JUNE 2004

Impact Factor: 1.5 · DOI: 10.1002/cem.877

JOURNAL OF CHEMOMETRICS
J. Chemometrics 2004; **18**: 341–349
Published online 31 December 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.877

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. A. Fernández Pierna[1], V. Baeten[1], A. Michotte Renier[1], R. P. Cogdill[2] and P. Dardenne[1]*

[1]Walloon Agricultural Research Centre (CRA-W), Quality of Agricultural Products Department, Chaussée de Namur 24, B-5030 Gembloux, Belgium
[2]Graduate School of Pharmaceutical Sciences, Duquesne University, 410 Mellon Hall, 600 Forbes Avenue, Pittsburgh, PA 15282, USA

Received 10 October 2003; Revised 3 August 2004; Accepted 1 September 2004

This study concerns the development of a new system to detect meat and bone meal (MBM) in compound feeds, which will be used to enforce legislation concerning feedstuffs enacted after the European mad cow crisis. Focal plane array near-infrared (NIR) imaging spectroscopy, which collects thousands of spatially resolved spectra in a massively parallel fashion, has been suggested as a more efficient alternative to the current methods, which are tedious and require significant expert human analysis. Chemometric classification strategies have been applied to automate the method and reduce the need for constant expert analysis of the data. In this work the performance of a new method for multivariate classification, support vector machines (SVM), was compared with that of two classical chemometric methods, partial least squares (PLS) and artificial neural networks (ANN), in classifying feed particles as either MBM or vegetal using the spectra from NIR images. While all three methods were able to effectively model the data, SVM was found to perform substantially better than PLS and ANN, exhibiting a much lower rate of false positive detection. Copyright © 2004 John Wiley & Sons, Ltd.

KEYWORDS: compound feeds; NIR; imaging spectroscopy; chemometrics; PLS; ANN; SVM

1. INTRODUCTION

Since the emergence of the mad cow crisis in Europe, and all its socio-economic consequences, European Union (EU) regulatory agencies have undertaken many legal measures to assure the safety and quality of feedstuffs. One of the most important decisions has been to 'totally ban processed animal proteins (meat and bone meal (MBM)) in feedstuffs destined to farm animals which are kept, fattened, or bred for the production of food' [1,2]. In order to combat fraud and accidental contamination, the effective enforcement of this regulation requires accurate and efficient analytical methods capable of analysing thousands of samples per year. Classical microscopy [3,4] was the first method available for the detection of meat and bone meal (MBM) in feedstuffs; this technique is reliable but tedious, since it requires visual observation and interpretation by an experienced analyst. Alternative techniques, mainly based on molecular biology (PCR and ELISA), chromatography (HPLC) and near-infrared spectroscopy (NIRS) [5], have been evaluated to help improve speed and reduce costs.

Near-infrared microscopy (NIRM) [6,7] was proposed as an alternative method; NIRM has an advantage in that each particle is evaluated based on its chemical properties rather than appearance, thus reducing human subjectivity. This technique was shown to perform well in discriminating the different ingredients found in compound feeds, including MBM. Whereas this method allows simultaneous detection of a larger number of ingredients in a single analysis, it is no faster than classical microscopy, since the spectra are collected in a serial (particle-by-particle) manner.

Recent developments in NIR focal plane array (FPA) technology offer a solution to this problem in the form of imaging spectroscopy, which combines the advantages of spectroscopic and microscopic methods along with much faster sample analysis, since the spectral data are acquired in parallel. An imaging spectrometer gathers spectral and spatial data simultaneously by recording sequential images of a predefined sample; each image plane is collected at a single wavelength band [8]. In the present study, reflectance images were collected in the 900–1700 nm window, with an

*Correspondence to: P. Dardenne, Walloon Agricultural Research Centre (CRA-W), Quality of Agricultural Products Department, Chaussée de Namur 24, B-5030 Gembloux, Belgium.
E-mail: dardenne@cra.wallonie.be

increment of 10 nm. The image planes are stacked to form a three-sided matrix, where the first two axes (x and y) define the image plane (field of view (FOV)) and the third axis (z) corresponds to the spectrum at each pixel location in the FOV. The imaging spectrometer utilizes an InGaAs FPA with 240×320 pixels (76 800 spectra per scan), along with a liquid crystal tuneable filter (LCTF) for wavelength selection. The effective FOV covers approximately 5 cm^2, allowing simultaneous analysis of 300–400 particles. Where applicable, spectral images are background corrected and converted to absorbance units prior to further analysis [8].

In order to locate features and extract and analyse the information from spectral image data, a combination of image processing and chemometric techniques may be applied [9]. Classical chemometric methods such as partial least squares (PLS) [10] and artificial neural networks (ANN) [11] are well-known, proven techniques for both classification and regression analysis of multivariate data such as NIR spectra. In this work a more recent technique, support vector machines (SVM) [12–26], was compared with traditional chemometric methods of classification and evaluated for future predictive performance in analysing compound feeds for MBM [27,28].

The choice of SVM as classification method is justified by the results obtained in References [14,18], where the great performance of SVM becomes evident. This performance can be partially explained by the uniqueness of the SVM solution [13] for the problems of pattern recognition and regression estimation.

2. THEORY

2.1. Support Vector Machines (SVM) for classification [12–26]

The objective of an SVM classifier is to derive a function $f(x)$ that describes the decision boundary or hyperplane which optimally separates two classes of points (Figure 1). For a linearly separable data set containing n NIR spectra $X = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ with $x_i \in \mathbb{R}^n, y_i \in \{-1, +1\}, i = 1, \ldots, n$, the hyperplanes are of the type $\langle w \cdot x \rangle + b = 0$, corresponding to decision functions

$$f(x) = \text{sign}(\langle w \cdot x \rangle + b) = \begin{cases} +1, \langle w \cdot x \rangle + b > 0 \\ -1, \langle w \cdot x \rangle + b \leq 0 \end{cases} \quad (1)$$

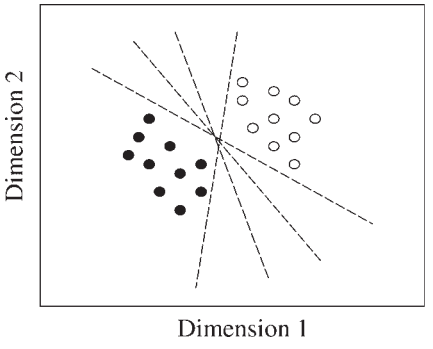

Figure 1. Various two-class, two-dimensional decision boundaries.

where b and w are the hyperplane parameters (offset and weight vector respectively) and $\langle w \cdot x \rangle$ represents the inner product of w and x.

During optimization, the SVM classifier locates the decision boundary with maximal margin among all possible hyperplanes. Margin is the distance d from the hyperplane to the closest point for both classes of points [12]: $M = \min_{1 \leq i \leq n} d(x_i, f(w, b)) = 2/\|w\|$. In order to maximize the margin, one has to minimize $\|w\|$ subject to some constraints:

$$\max_{w,b} \min_{1 \leq i \leq n} d(x_i, f(w, b)) = \begin{cases} \min \frac{1}{2} \|w\|^2 \\ \forall i, y_i(\langle w \cdot x_i \rangle + b) \geq 1 \end{cases} \quad (2)$$

This is a constrained quadratic optimization problem that can be solved by using the method of Lagrange multipliers [13]:

$$\begin{cases} L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\langle w \cdot x_i \rangle + b) - 1] \\ \forall i, \alpha_i \geq 1 \end{cases} \quad (3)$$

where the α_i are non-negative parameters learned from the data (Lagrange multipliers). This is a major advantage of the SVM framework over back-propagation ANN, since optimization is a determinate operation where there is only one minimum solution [26] (no problems with local minima). This equation can be transformed to its dual problem under the Karush–Kuhn–Tucker conditions [12,29]:

$$\begin{cases} \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle \\ \forall i, \alpha_i \geq 0 \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases} \quad (4)$$

The solution of this dual formulation reduces to

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i \quad (5)$$

Each point x_i is described by a Lagrange multiplier. Those points with $\alpha_i [y_i(\langle w \cdot x_i \rangle + b) - 1] = 0$ will have non-zero Lagrange multipliers ($\alpha_i > 0$) and are referred to as the 'support vectors'. All other samples may simply be omitted when the model is applied to new data, since they will bear no influence on prediction, which reduces the computational overhead. Thus the optimal decision function can be described as

$$f(x) = \text{sign}\left(\sum_{i=1}^{n_{sv}} \alpha_i y_i \langle x_i^{sv} \cdot x \rangle + b \right) \quad (6)$$

where n_{sv} represents the number of support vectors.

In the case of non-linearly separable data it is impossible to find a linear decision boundary that will perfectly classify all the training samples. Thus a trade-off must be made between maximizing the margin while minimizing the number of misclassified training samples. This is accomplished by adding a user-selected regularization parameter C [14]:

$$\begin{cases} \min \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \right) \\ \forall i, y_i(\langle w' \cdot x_i \rangle + b) \geq 1 - \xi_i \end{cases} \quad (7)$$

where $\xi_i = |1 - y_i f(x_i)|$ is a non-negative slack variable which estimates classification error (relative to the decision function) at point (x_i, y_i). Errors are found when $\xi_i > 1$.

The first term applies a penalty to the magnitude of weight vector, which seeks to limit complexity (and error of generalization) and maximize the margin; the second term forces the optimization towards minimal error. Thus the regularization parameter C allows the user to adjust the trade-off between error minimization and maximal margin estimation. As before, Equation (7) is solved by using the method of Lagrange multipliers, arriving at the same solution as in the case of linearly separable data (Equation (2)), but with an additional constraint $0 \leq \alpha_i \leq C$.

While it is impossible to completely separate non-linearly separable classes with the linear hyperplanes described in the previous equations, it is often possible to map the training data to a higher-dimensional feature space Φ, where the data $\Phi(x)$ may be separated with a higher-dimensional linear hyperplane [15]:

$$\Phi\colon \mathbb{R}^d \rightarrow \mathbb{R}^D (D \gg d)$$
$$x \rightarrow \Phi(x)$$

In that case the decision function becomes

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{n_{SV}} \alpha_i y_i \langle \Phi(x_i) \cdot \Phi(x) \rangle + b\right) \tag{8}$$

$$0 \leq \alpha_i \leq C$$

The relationship between two vectors in the feature space Φ is efficiently represented by defining a kernel function $k(x_i, x_j)$, where

$$k(x_i, x_j) = \langle \Phi(x_i) \cdot \Phi(x_j) \rangle \tag{9}$$

It is only through the choice of a non-linear kernel function that non-linear models can be formed within the SVM framework. However, as long as some necessary conditions are met (Mercer conditions) [16], any one of many available kernel functions can be used, e.g. linear, Gaussian radial basis function (RBF), polynomial, sigmoid, etc. As explained by Chih-Wei *et al*. [17], RBF is the most reasonable choice because of its simplicity and ability to model data of arbitrary complexity. They proved that the linear kernel is a special case of RBF. Another reason for using RBF is the number of hyperparameters, which influences the complexity of model selection (polynomial kernel, for instance, has more hyperparameters than RBF). They show that the RBF kernel is computationally easier than, for instance, the polynomial one, where the kernel value may go to infinity or zero when the degree increases. As proved also by Vapnik [15], the sigmoid kernel does not fulfil the necessary conditions under some parameters. The equation of the RBF kernel is as follows:

$$k(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{10}$$

where σ defines the width of the Gaussian function and can be used to adjust the degree of generalization. Thus the separating hyperplane is constructed as

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{n_{SV}} \alpha_i y_i k(x_i, x) + b\right) \tag{11}$$

$$0 \leq \alpha_i \leq C$$

Along with C, the implementation of Equation (11) with an RBF kernel function requires the user to specify an additional parameter σ. The correct selection of both parameters is critical to SVM predictive performance. While a number of optimization schemes have been proposed in the literature [18], no consensus has been found. For this work the proposal of Chih-Wei *et al*. [17] was adopted, employing a 'grid search' on C and σ where the parameter combination with the best cross-validation accuracy is selected.

3. EXPERIMENTAL

3.1. Software

All computations, chemometric analyses and graphics were executed with programs developed in Matlab v6.5 (The Mathworks, Inc., Natick, MA, USA). PLS calibrations were derived using the SIMPLS algorithm [30] included in the PLS Toolbox (Eigenvector Research, Inc., Manson, WA, USA), and ANN calibrations were applied using the back-propagation procedure [11]. Different algorithms have been proposed in the literature to perform SVM for classification [19–21]. Here the Lin's Lib SVM v2.33 algorithm was used [22].

3.2. Model construction

In all cases the chemometric algorithms tested were used to construct models to classify the origin of feed particles as either 'animal' or 'vegetal' in order to discriminate between MBM particles (not allowed) and plant particles (allowed).

Training (calibration) spectra were drawn from two groups: spectra coming from 26 pure animal meals as the first group, and spectra coming from 59 pure vegetal meals as the second group. Each spectrum corresponded to a pixel occupied by a particle included in the FOV of the NIR imaging spectrometer. The animal and vegetal materials analysed were selected to span the diversity of materials mainly used for the formulation of compound feeds. These samples were selected from the sample bank existing at the CRA-W [7] and from the European sample bank constructed in the framework of the STRATFEED project [31]: 'Development and validation of methods for the detection and quantification of mammalian tissues in feedingstuffs'.

In total, more than 267 000 spectra have been collected from pure animal and vegetal meals; each spectrum spans the 900–1700 nm range, in 10 nm increments. Owing to limitations in storage space, memory and calculation capacity, and since each particle occupied multiple pixels in the FOV during scanning, the mean spectrum was drawn from each particle for calibration, which reduced the training set to 5521 spectra, i.e. 2233 animal and 3288 vegetal particles. During preliminary testing, a number of different preprocessing methods were tested for their effect on accuracy, but no real improvement was found in any case. Thus all spectra were kept as raw absorbance units. A dummy variable was attributed to represent each group (1 for animal and -1 for vegetal) and was considered as reference value (y) during the derivation of the PLS, ANN and SVM classifiers. Each classification method was trained independently, with the best model of each type selected for the comparison and described below.

3.2.1. PLS

The RMSE (root mean square error) was determined by leave-one-out cross-validation. The optimal model complexity is selected as the one with the minimum RMSE. The best PLS model required 15 factors and achieved an RMSE of 0.397.

3.2.2. ANN

The training data set was split using the Kennard–Stone method [32] into three subsets, (1) a calibration set, (2) a monitoring set and (3) a test set, in order to determine the optimal topology of the network. Separate monitoring and test sets are required for ANN derivation to avoid overfitting of the training data. The optimal topology was found to be [15:8:1], i.e. 15 input nodes, a single hidden layer with eight nodes (hyperbolic tangent transfer function), and one output node (linear transfer function). This ANN model produced an RMSE of 0.139.

3.2.3. SVM

Different kernels have been tested on these data, and the results showed that the best choice, as concluded by Chih-Wei *et al*. [17], is the RBF kernel. The SVM parameters were optimized using the 'grid search' method [17] with a fixed calibration and validation set, split from the training data using the Kennard–Stone method. The optimal parameter settings for C and σ are then selected as the values that give the minimum RMSE and the maximum correct classification rate. During optimization of the SVM parameters, it was observed that: (i) when C is increased, the second term of Equation (7) dominates, forcing SVM towards a solution with the least training error, which decreases the amount of regularization, causing more background pixels (neither animal nor vegetal) to be classified as animal; (ii) when C is decreased, more emphasis is placed on reducing w, thus maximizing the margin and emphasizing regularization. Consequently, a larger number of calibration samples are retained as support vectors, which increases the computation time of prediction. In the second case, animal particles begin to be classified as vegetal, but no plant particles are misclassified. This feature is very attractive, since the overall objective of this study is to develop a method for use by regulatory laboratories. While it is important that both MBM and vegetal samples are classified correctly, more importance is given to the correct identification of vegetal particles as free of MBM for the following reasons.

1. A false detection of MBM is unfair and can severely damage the reputation of honest and scrupulous processors and manufacturers.
2. Since the testing of a single feed sample involves the analysis of several hundred, perhaps thousands of individual particles, it is unlikely that all MBM particles in a truly contaminated sample will be misclassified as vegetal. Even a single MBM particle detection could signal more exhaustive analysis of the feed sample in question.
3. Constant human intervention to verify false MBM detections would be very laborious and expensive, while the simple threat of a credible feed testing system can influence the methods of would-be offending processors and manufacturers.

The optimal parameter settings were found to be $C = 100$ and $\sigma = 10$; with these parameter settings, SVM performed with an RMSE of 0.102.

3.3. Classification of a new data set

An independent validation data set was gathered (the 'cross' data set) for rigorous comparison of the predictive performance of each chemometric algorithm tested. This 'cross' data set, which was created in the same manner as the calibration samples, using the same imaging instrument, consists of 76 800 spectra, which corresponds to 211 particles placed on the sample presentation device. Figure 2 is an image of the 'cross' data set at 1500 nm which shows the distribution of the 211 particles within the FOV. As indicated in the figure, the data consist of several pure animal (MBM) particles (49) arranged in the middle of plant particles (162) in the shape of a cross.

The prediction of the 'cross' data was handled in two ways. In the first method the mean spectrum of each particle was extracted (as was done for the calibration set) and

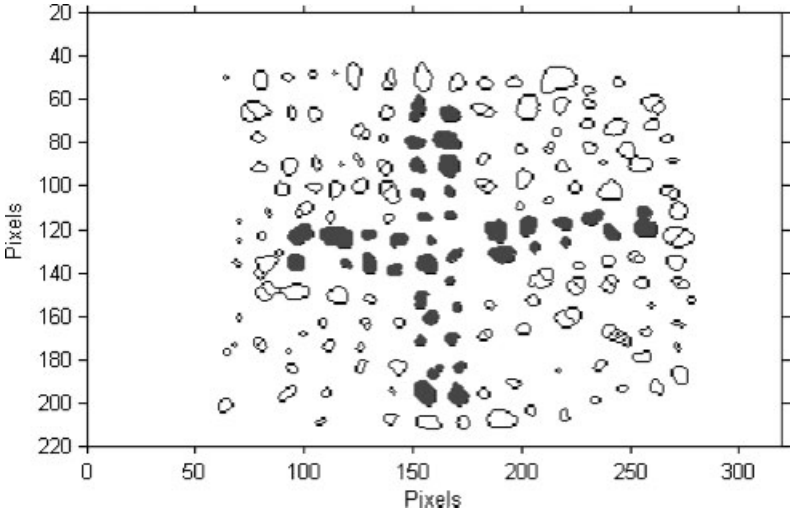

Figure 2. Image obtained with the imaging spectrometer at 1500 nm for the 'cross' data set. The animal particles are indicated in black.

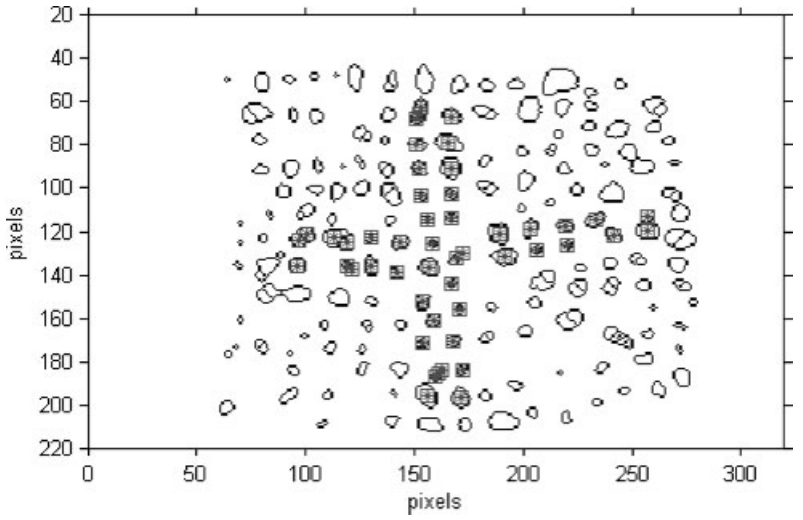

Figure 3. Results of the PLS 'animal versus vegetal' model (15 LVs) for the 'cross' data set using the mean of each particle. Black dots are classified as animal.

predictions were derived for the mean spectra. In the second method the classification models were applied to all the pixels in the FOV of the 'cross' data set image (76 800 spectra). This second method has two important consequences for the aim of this study: first, it is important in order to obtain a quantification of the possible contamination of the data set; second, it will be useful in order to simplify the spreading procedure by proving that in the future the particles can be arranged on the holder in a thin layer rather than by carefully placing each particle in a distinct position.

The results of the first test, where only the mean spectrum of each particle is used, are shown in Figures 3–5 for PLS, ANN and SVM respectively. For the sake of clarity, the true perimeter of each particle has been traced in black using the 'watershed' segmentation method included in the Matlab Image Processing Toolbox. The best results were obtained working with PLS, but, as can be seen in the images, all three techniques arrive at nearly the same prediction result when mean particle spectra are used for prediction. In all cases, most of the animal particles are well detected and no vegetal particles are misclassified.

In the second test the same calibration models were applied to derive predictions from all 76 800 spectra of the 'cross' data set. In this case the image contains pixels from animal and vegetal particles as well as background pixels from the sample presentation device, which were not included in the training data set. Figures 6–8 illustrate results for the PLS, ANN and SVM models respectively (pixels classified as MBM are grey, while pixels classified as vegetal are white). For the PLS model, although the models generally classified MBM particles correctly, an abnormally high percentage of background particles were classified incorrectly as animal. This would suggest that the linear decision boundary derived by PLS is unable to sufficiently generalize during prediction of the background pixels, which were not represented in the training data and may

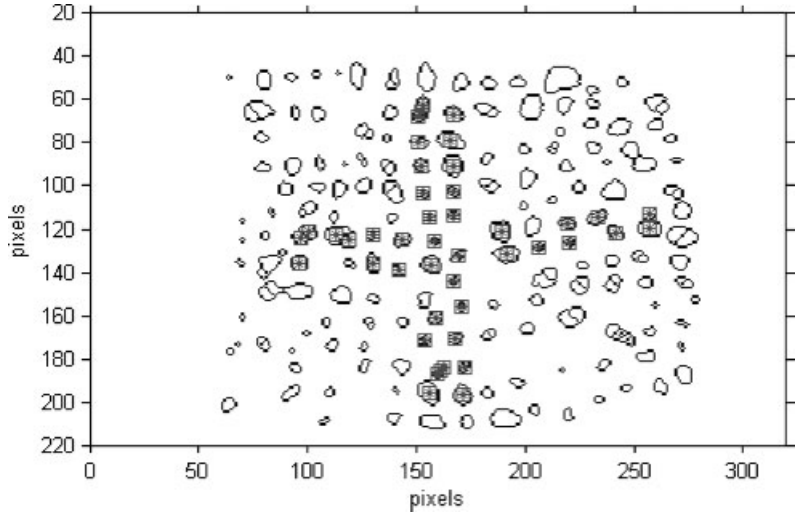

Figure 4. Results of the ANN 'animal versus vegetal' model (15:8:1) for the 'cross' data set using the mean of each particle. Black dots are classified as animal.

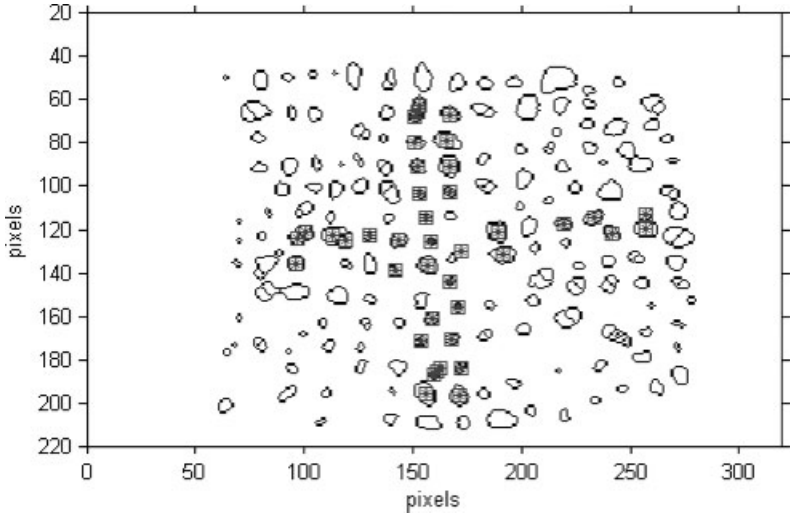

Figure 5. Results of the SVM model 'animal versus plant' ($C = 100$, $\sigma = 10$) for the 'cross' data set using the mean of each particle. Black dots are classified as animal.

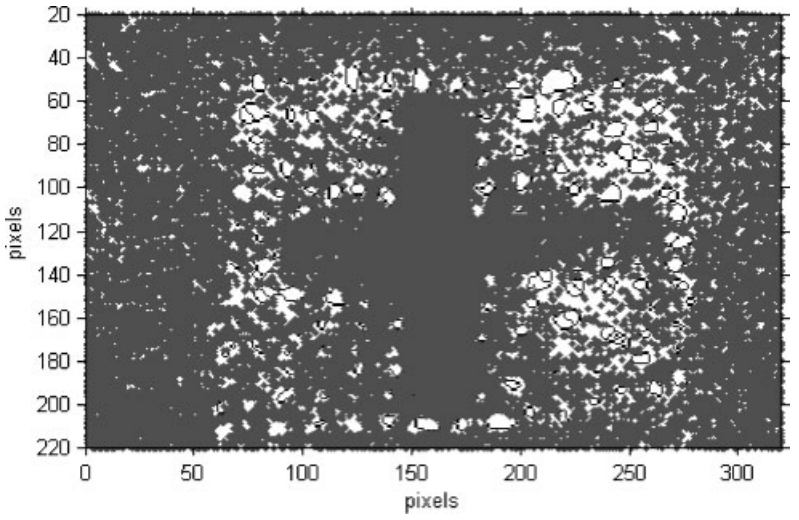

Figure 6. Results of the PLS model for the 'cross' data set using the whole spectra (76 800 pixels). Black dots are classified as animal.

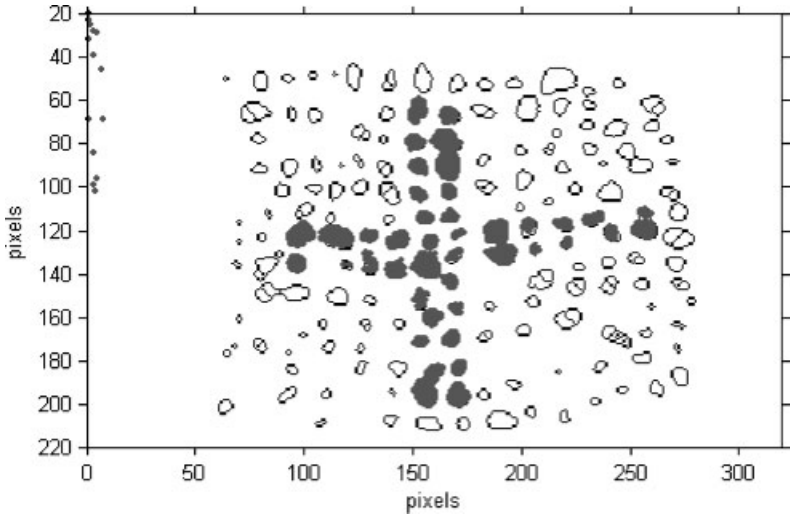

Figure 7. Results of the ANN model for the 'cross' data set using the whole spectra (76 800 pixels). Black dots are classified as animal.

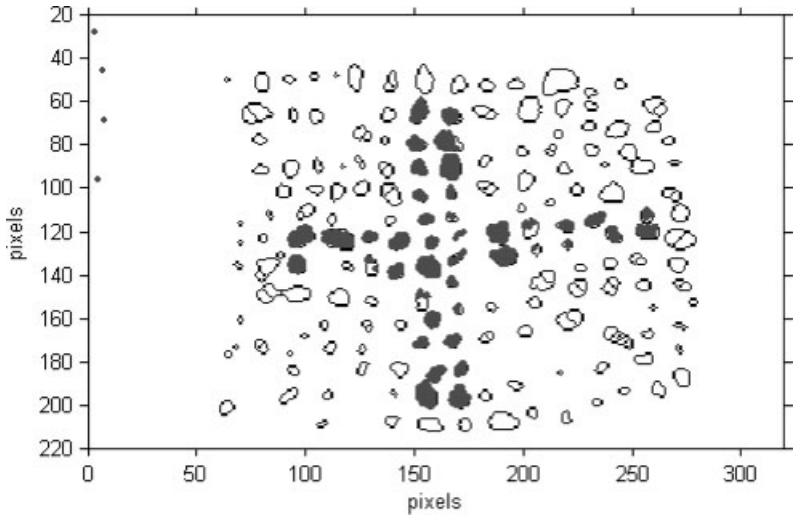

Figure 8. Results of the SVM model for the 'cross' data set using the whole spectra (76 800 pixels). Black dots are classified as animal.

not be linearly separable from the MBM or vegetal spectra. For ANN and SVM the detected animal particles corresponded better with the true MBM particles (see Figure 2), and virtually no vegetal or background particles were misclassified. By utilizing the watershed mask, the number of particles and their corresponding pixels can be quantified, which allows estimation of the percentage of correctly classified pixels.

Table I contains the results (in pixels) for the PLS, ANN and SVM 'animal versus vegetal' models. For PLS, 99.7% of true MBM pixels were correctly classified as MBM; the ANN and SVM models achieved 93.4% and 81.6% correct classification respectively. At first glance the initial conclusion would be that PLS produced better classification results than the other techniques. However, neither the ANN nor SVM model misclassified any of the vegetal pixels as MBM, compared with 2.08% misclassified for PLS. Moreover, the PLS results further deteriorate when the classification performance of background pixels is considered (Table I and Figure 7). More than 46% of the background pixels were misclassified as MBM by the PLS model, while only 1.09% and 0.35% were misclassified by the ANN and SVM models respectively. Thus, for PLS, even if all the animal particles were correctly detected, the reliability of detection would still be low, since it is quite probable that any non-vegetal pixels will be classified as MBM whether they are indeed animal or not. This is an important point, because one of the aims of our work is to be able to simplify the spreading procedure in order to reduce the time of analysis. Till now the particles were separated manually on the holder, and the methodology of the mean spectra seems to be good enough in that case. However, when the particles are not separated, i.e. disposed in a thin layer, the creation of this mask becomes impossible and so we are obliged to work with all the pixels. Furthermore, since the origination and composition of compound feeds may vary drastically, it must be assumed that periodically a new, non-mammalian, feed compound (not included in the training data set) will be encountered. Based on the observed results, there is little confidence that a PLS model would correctly classify the new compound. This would not likely be the case for ANN and SVM, where, even if some MBM particles are not

Table I. Results for the 'cross' data set using the PLS, ANN and SVM 'animal versus vegetal' models

		Detected animal		
Source	*n*	PLS	ANN	SVM
Animal	49 particles			
(*a*)	1099 pixels	1096 pixels	1027 pixels	897 pixels
		99.73%	93.45%	81.62%
Plant	162 particles			
(*b*)	2643 pixels	55 pixels	0 pixels	0 pixels
		2.08%	0.00%	0.00%
Holder				
(*c*)	73 058 pixels	33 946 pixels	793 pixels	257 pixels
		46.46%	1.09%	0.35%
Percentage of pixels correctly classified as animal ($a/(a+b+c)$)		3.12%	56.43%	77.73%

detected, there is sufficient confidence in the rate of false detection that the contaminated sample would be correctly identified and 'clean' samples would more likely pass analysis with fewer false positive MBM detections, even if some new feed ingredient were encountered. In order to see if the differences between the performances of the methods are significant, a McNemar test [33] with the continuity correction was applied in order to calculate the two-tailed *P* value. This test shows that SVM performs significantly better than PLS and equally as well as ANN. In the case of PLS the *P* value was less than 0.0001, so by conventional criteria this difference is considered to be extremely statistically significant. In the case of ANN the *P* value was 1.000, showing that the difference is not statistically significant. The last row of Table I shows the percentage of pixels correctly classified as animal by each of the techniques, i.e. the ratio between the animal pixels classified as animal and the total pixels (animal, plant and background) classified also as animal. These results clearly show the advantage of SVM over the other methods during prediction of independent samples.

4. CONCLUSION

This work was carried out in an effort to improve the enforcement of legislation banning the use of meat and bone meal in compound animal feeds. The methodology developed herein effectively combines imaging spectroscopy and chemometrics. The integrated system is able to rapidly analyse many samples using either PLS, ANN or SVM classification models.

While all three chemometric classification algorithms tested performed admirably in analysing the training data, and when analysing the data of spectral types which were represented within the training data set (vegetal and MBM particles), the ANN and SVM models showed superiority over PLS in generalization ability when unmodelled data were encountered (i.e. background pixels).

Even if the differences between the performances of SVM and ANN are not significant, SVM is preferred to ANN for two reasons. The first reason is that SVM gives fewer false positive than ANN. This is crucial for the development of methods aiming to support legislative decisions stipulating a total ban (0%). The second reason is that, as explained by Gunn [12], traditional neural network approaches have suffered difficulties with generalization, producing models that can overfit the data. This is due to the large number of model parameters to be optimized. As with all other methods, ANN optimizes a quadratic (or other type of) criterion, and the best model is obtained when this optimal value is found. For ANN one has to train for a long time to find this, and usually it gives less precise predictions than stopping the training earlier. This is due to the large number of regression parameters that need to be fitted. ANN does not lead to one global (unique) solution owing to differences in the initial weight set (neural networks have to be trained with randomly chosen initial weight settings). SVM is trained as a convex optimization problem resulting in a global solution (no local minima can occur), which in many cases yields unique solutions, i.e. the solution is found in a deterministic manner. Owing to this specific optimization procedure, it is assured that overtraining is avoided and the SVM solution is general. Additionally, the SVM decision function can be easily evaluated owing to the reduced number of training data that contribute to the solution (i.e. the support vectors).

With this study it was proved that the combination of NIR imaging spectroscopy and certain non-linear chemometric classification techniques should allow a regulatory laboratory to certify and quantify the presence of meal and bone meal in common samples of processed animal feed.

REFERENCES

1. European Commission (EC). Council Decision 2000/776/EC of 4 December 2000 concerning certain protection measures with regard to bovine spongiform encephalopathy and the feeding of animal protein. *Off. J. Eur. Commun.* 2000; **L306**: 32–33.
2. European Commission (EC). Regulation (EC) No. 1774/2000 of the European Parliament and the Council of 3 October 2002 laying down health rules concerning animal by-products not intended for human consumption. *Off. J. Eur. Commun.* 2002; **L273**: 1–95.
3. Gizzi G, van Raamsdonk LWD, Baeten V, Murray I, Berben G, Brambilla G, von Holst C. In *An Overview of Tests for Animal Tissues in Animal Feeds Used in the Public Health Response Against BSE*, Lasmezas CI, Adams DB (eds). OIE Scientific and Technical Review 2003; **22**(1): 311–331.
4. Frick G, Rietschi A, Hauswirth H. Mikroskopische *Untersuchung von Futtermitteln*. AGRARForschung 2002; **9**(11–12): 497–504.
5. Baeten V, Dardenne P. *The Contribution of Near Infrared Spectroscopy to the Fight Against the Mad Cow Epidemic*. NIR News, 2001; **12**(6).
6. Piraux F, Dardenne P. *Feed Authentication by Near-Infrared Microscopy, Proceedings of the 9th International Conference NIR-99* 1999, Verona, Italy; 535–542.
7. Baeten V, Michotte-Renier A, Sinnaeve G, Dardenne P. *Analyse of Feedingstuffs by Near-infrared Microscopy (NIRM): Detection and Quantification of Meat and Bone Meal (MBM), Proceedings of the Sixth Food Authenticity and Safety International Symposium (FASIS)* 2001, Nantes, France; 1–11.
8. The Spectral dimensions Website. http://www.spectral-dimensions.com [Accessed on 30 July 2004].
9. Harwood J, Aparicio R. *Handbook of Olive Oil*. Aspen: Gaithersburg, MD, 2000; chaps 10 and 14.
10. Martens H, Naes T. *Multivariate Calibration* (2nd edn), vol. 1. Wiley: Chichester, 1989.
11. Despagne F, Massart DL. Tutorial review: neural networks in multivariate calibration. *The Analyst* 1998; **123**: 157R–178R.
12. Gunn SR. *Support Vector Machines for Classification and Regression*. Technical report, Image speech and intelligent systems research group, University of Southampton, UK, 1998. Available on http://www.ecs.soton.ac.uk/~srg/publications/pdf/SVM.pdf [Accessed on 30 July 2004].
13. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 1998; **2**(2): 121–167.
14. Belousov AI, Verzakov SA, von Frese J. *Support Vector Machines: A Versatile and Powerful Approach to Data Analysis, Poster at the Gordon Conf. on Statistics and Chemical Engineering*, Williamstown, MA, 2001. Available on http://www.amstat.org./sections/spes/GRC2001.htm [Accessed on 30 July 2004].
15. Vapnik VN. *The Nature of Statistical Learning Theory* (2nd edn). Springer: New York, 2000.

16. Cristiani N, Shawe-Taylor J. *An Introduction to Support Vector Machines*. Cambridge University Press: Cambridge, 2000.
17. Chih-Wei H, Chih-Chung C, Chih-Jen L. *A Practical Guide for Support Vector Classification*. Available on http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf [Accessed on 30 July 2004].
18. Belousov AI, Verzakov SA, von Frese J. Applicational aspects of support vector machines. *J. Chemometrics* 2002; **16**: 482–489.
19. The Kernel Machines Website. http://www.kernel-machines.org [Accessed on 30 July 2004].
20. Image speech and intelligent systems research group. University of Southampton, UK, 1998. Available on http://www.isis.ecs.soton.ac.uk/isystems/kernel [Accessed on 30 July 2004].
21. Suykens JAK, van Gestel T, De Brabanter J, De Moor B, Vandewalle J. *Least Squares Support Vector Machines*. World Scientific: Singapore, 2002.
22. Chih-Chung C, Chin-Jen L. National Taiwan University, 2002. Available on http://www.csie.ntu.edu.tw/~cjlin/libsvm [Accessed on 30 July 2004].
23. Evgeniou T, Pontil M, Papageorgiou C, Poggio T. *Image Representations for Object Detection Using Kernel Classifiers, Lecture at the Fourth Asian Conference on Computer Vision, ACCV* 2000, Taipei.
24. Cogdill RP, Dardenne P. Least-squares support vector machines for chemometrics: an introduction and evaluation. *J. Near Infrared Spectrosc.* 2004; **12**(2): 93–100.
25. Evgeniou T, Pontil M. *Workshop on Support Vector Machines: Theory and Applications, Lecture at the Advanced Course on Artificial Intelligence (ACAI 99)*, 1999, Chania, Greece.
26. Cornuéjols A. Bulletin de l'AFIA, 2002; 51.
27. Koehler FW, Lee E, Kidder LH, Lewis EN. Near-infrared chemical imaging: the practical choice. *Spectroscopy Europe* 2002; **41**: 3.
28. Baeten V, Dardenne P. Spectroscopy: developments in instrumentation and analysis. *Grasas y Aceites* 2002; **53**(1): 45–63.
29. Kuhn HW, Tucker AW. Nonlinear Programming. In *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics*. University of California Press: Berkeley, CA, 1951; 481–492.
30. De Jong S. SIMPLS: an alternative approach to partial least squares regression. *Chemome. Intell. Lab. Syst*. 1993; **18**: 251–263.
31. Website. http://stratfeed.cra.wallonie. be/ [Accessed on 30 July 2004].
32. Kennard RW, Stone LA. Computer aided design of experiments. *Technometrics* 1969; **11**(1): 137–149.
33. Agresti A. *Categorical Data Analysis*. Wiley: New York, 1990.