Validation Using Sensitivity and Target Transform Factor Analyses of Neural Network Models for Classifying Bacteria from Mass Spectra

Peter de B. Harrington
Center for Intelligent Chemical Instrumentation, Department of Chemistry and Biochemistry, Clippinger Laboratories, Ohio University, Athens, Ohio, USA

Kent J. Voorhees, Franco Basile, and Alan D. Hendricker
Chemistry and Geochemistry Department, Colorado School of Mines, Golden, Colorado, USA

Temperature constrained cascade correlation networks (TCCCNs) are computational neural networks that configure their own architecture, train rapidly, and give reproducible prediction results. TCCCN classification models were built using the Latin-partition method for five classes of pathogenic bacteria. Neural networks are problematic in that the relationships among the inputs (i.e., mass spectra) and the outputs (i.e., the bacterial identities) are not apparent. In this study, neural network models were constructed that successfully classified the targeted bacteria and the classification model was validated using sensitivity and target transformation factor analysis (TTFA). Without validation of the classification model, it is impossible to ascertain whether the bacteria are classified by peaks in the mass spectrum that have no causal relationships with the bacteria, but instead randomly correlate with the bacterial classes. Multiple single output network models did not offer any benefits when compared to single network models that had multiple outputs. A multiple output TCCCN model achieved classification accuracies of $96 \pm 2\%$ and exhibited improved performance over multiple single output TCCCN models. Chemical ionization mass spectra were obtained from in situ thermal hydrolysis methylation of freeze-dried bacteria. Mass spectral peaks that pertain to the neural network classification model of the pathogenic bacterial classes were obtained by sensitivity analysis. A significant number of mass spectral peaks that had high sensitivity corresponded to known biomarkers, which is the first time that the significant peaks used by a neural network model to classify mass spectra have been divulged. Furthermore, TTFA furnishes a useful visual target as to which peaks in the mass spectrum correlate with the bacterial identities.   (J Am Soc Mass Spectrom 2002, 13, 10–21) © 2002 American Society for Mass Spectrometry

Mass spectra of microorganisms are complex and often require pattern recognition to overcome variations that arise from the measurement and biological factors. Artificial neural networks (ANNs) have been successfully applied to these spectra to characterize and identify microorganisms such as bacteria [1–3]. In general, ANNs have been shown to be a rapid and accurate method for classification and discrimination of various microorganisms using pyrolysis mass spectrometry (Py-MS) data. Goodacre et al. have demonstrated that ANNs can be utilized to discriminate between the mass spectra from methicillin-resistant and methicillin-susceptive *Staphylococcus aureus* [4] and to rapidly identify urinary tract infection bacteria [5]. Sisson et al. [6] have achieved rapid detection of verocytotoxin production status in *Escherichia coli* by analysis of Py-MS data with ANN models. Freeman et al. [7] used ANN models for resolution of batch variation in bacteria. Kenyon et al. [8] have applied ANNs to identify members of a S*treptomycete* species-group.

The deployment of biological weapons by terrorists concerns a number of governments. Pathogenic bacteria are potential agents of mass destruction and are relatively inexpensive to produce. Researchers at the Colorado School of Mines [9–11] have demonstrated that thermal hydrolysis methylation-mass spectrometry

© 2002 American Society for Mass Spectrometry. Published by Elsevier Science Inc.
1044-0305/02/$20.00
PII S1044-0305(01)00345-2

Received April 26, 2001
Revised October 11, 2001
Accepted October 11, 2001

(THM-MS) coupled with pattern recognition techniques can provide sensitive detection of targeted bacteria in various environments. By combining sensitive, field-portable pyrolysis mass spectrometry with neural network pattern recognition software, rapid detection and classification of pathogenic bacteria may be obtainable.

The bacteria and spores were chosen for this study because they are potential biological agents. *Bacillus anthracis* was studied in both vegetative and sporulated states. *Bacillus anthracis* is a Gram-positive bacterium which causes the disease anthrax. The other three bacteria were Gram-negative. *Yersinia pestis* is responsible for the plague, *Francisella tularensis* causes tularemia, which is also known as rabbit fever, and *Brucella melitensis* causes brucellosis, which is also known as Gibraltar, Malta, Rock, and undulant fever.

Theory

Artificial neural networks are powerful tools for developing classifiers that can relate mass spectra to user-specified properties. In this work, the user-specified property is the bacterial class. The neural network builds a model that allows the bacteria to be classified by their mass spectra. The problem with neural network classification is that it is difficult to verify if the model is classifying the spectra on biomarker peaks or artifacts that correlate with the experiment. For example, if samples from a bacterium differed from the other bacterial samples in that they contained a contamination, the classification model may perform well in recognizing the bacterium. However, the model may be recognizing the bacterium by peaks that arise from the contamination instead of peaks that arise from characteristic bacterial components (i.e., biomarkers).

TCCCNs can rapidly construct robust classification models. The characteristic properties of the TCCCN will be summarized, but detailed papers on the TCCCN can be found in the literature [12–14]. The cascade correlation network (CCN) trains much faster than other artificial neural networks because it adjusts only a single processing unit at a time [15]. The hidden processing units are incrementally added to the network until the training error criterion is satisfied. The TCCCN is a self-configuring network that optimizes its architecture (i.e., number of layers and hidden units) during the model building process. With the cascade architecture, computational temperature may be optimized for each hidden unit [12]. The temperature parameter controls the slope of the sigmoid of each hidden unit's output so that selectivity is optimized [16]. Constraining the computational temperature allows the neural network model to avoid local minima during training. Unconstrained hidden units usually furnish crisp or binary logic for which the output values for the objects are extrema, e.g., either 0 or 1. The temperature constrained hidden units furnish fuzzy logic that characterizes uncertainty. As a result, TCCCN models are better at

interpolation among the training spectra than other types of neural networks.

The neural network outputs ranged between 0 and 1. A classification criterion of 0.5 was used, so that outputs that were greater than or equal to 0.5 indicated class membership.

We advocate constructing classification models to a predetermined relative error. Instead of training neural network models so that an external prediction set of data furnishes a minimum error, the TCCCN models were trained to achieve a 10% relative root mean square error of calibration (RRMSEC). This error is defined as

$$\text{RRMSEC} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}(\hat{y}_{ij} - y_{ij})^2}{\sum_{i=1}^{n}\sum_{j=1}^{p}(y_{ij} - \bar{y}_j)^2}} \tag{1}$$

for which n is the number of training objects (i.e., spectra), p is the number of classes, y_{ij} is the target value for object i and class j, $\hat{y}_{ij}$ is the network model output for spectrum i and class j, and $\bar{y}_j$ is the average target value for class j. Matrices are indicated by upper-case bold typeface and vectors as lower-case bold typeface. Each row of **Y** corresponds to a mass spectrum and comprises values of zero except for the component that designates the class. This component has a value of unity. The error is relative to the pooled standard deviation for each column or class of target output values. The relative root mean square error of prediction (RRMSEP) is a similar figure of merit except the target values and network outputs are obtained from the external prediction sets of data. Using relative standard errors allows the comparison of errors among different neural network models and applications.

The importance of each mass spectral peak to the neural network models can be assessed with sensitivity analysis, which has been described in detail [17]. The sensitivity of a neural network model is the gradient of a neural network model output with respect to the input variables. The neural network model is a nonlinear function F acting on an object $\mathbf{x}_i$ to estimate a set of output values $\hat{\mathbf{y}}_i$.

$$\hat{\mathbf{y}}_i = F(\mathbf{x}_i) \tag{2}$$

The spectrum $\mathbf{x}_i$, which has v variables, generates the estimate $\hat{\mathbf{y}}_i$ that has p components (i.e., one for each class or output unit). The sensitivity spectrum for class j and object i, $\mathbf{s}_{ij}$ is defined as

$$\mathbf{s}_{ij} = \nabla_j F(\mathbf{x}_i) \tag{3}$$

for which the gradient ∇ of the neural network model $F(\mathbf{x}_i)$ is calculated for output unit j that corresponds to class j. For calculating the gradient, an increment is used that is 0.1% of the maximum intensity of the mean

spectrum from class j. In this paper, the training objects are used, but prediction data sets can also be used. The average sensitivity for each bacterial class is reported.

Because the bacteria spectra were inherently complex, a method was needed to determine which features in the spectra correspond to each class of bacteria and validate the sensitivity spectra. Target transformation factor analysis (TTFA) is a useful visualization tool for complex multivariate data [18, 19]. TTFA furnishes factor spectra from the mass spectral data sets that characterize the differences among the mass spectra for the target classes of bacteria. Therefore, sensitivity and TTFA spectra should correspond to some extent. The TTFA spectra give the mass spectral features that correlate with each bacterial class, whereas the TCCCN sensitivity spectra give mass spectral features that discern each class. Another difference is that TTFA spectra depict only linear relations between the mass spectra and the classes, while the sensitivity spectra may also depict nonlinear relationships among the spectra and classes.

The first step of TTFA is to calculate the principal components from the mass spectral data. The principal components are combined and obliquely rotated so that one factor is obtained for each class of bacteria. The mass spectra for each class, when projected on its corresponding factor spectrum, will yield a score close to unity and yield a score close to zero on the factors that correspond to the other bacteria. The eigenvalues correspond to the relative amount of variance spanned by each principal component or factor spectrum.

Each mass spectrum was normalized to unit vector length to correct for any differences in concentration. The average spectrum of the mass spectral data set was subtracted from each spectrum, so that the constant features in the spectra would be removed. This set of mass spectral data was then decomposed using singular value decomposition (SVD) [20]. The decomposition yielded sets of row ($\mathbf{V}$) and column ($\mathbf{U}$) space eigenvectors and a square diagonal matrix of singular values ($\mathbf{S}$).

$$\mathbf{X} = \mathbf{USV}^{\mathrm{T}} \tag{4}$$

for which $\mathbf{X}$ is the set of normalized and mean-centered mass spectra. SVD gives the principal components for which $\mathbf{V}$ corresponds to the principal components and the mass spectral scores comprise the product of $\mathbf{US}$. The squared singular values characterize the relative amount of variance spanned by each principal component. Twenty-six of these components spanned 99% of the cumulative variance of the data set and they were used in the following factor analysis.

The matrix of class target values, $\mathbf{Y}$ as defined in eq 1, is also used as the target for factor rotation. The transformation matrix ($\mathbf{T}$) is used to rotate the row eigenvectors $\mathbf{V}$ to furnish factor spectra for each target class. The transformation matrix ($\mathbf{T}$) was calculated using the column eigenvectors from the SVD of the mass spectral data set $\mathbf{X}$ as given in eq 4.

$$\mathbf{T} = \mathbf{U}^{\mathrm{T}}\mathbf{Y} \tag{5}$$

$$\hat{\mathbf{Y}} = \mathbf{UT} \tag{6}$$

for which $\mathbf{Y}$ is composed of columns of target values that were used to train the neural networks, $\mathbf{T}$ is the transformation matrix, $\mathbf{U}$ is the set of principal (i.e., reduced set of 26) column-space eigenvectors, and $\hat{\mathbf{Y}}$ is the estimated target matrix. The factor spectra, $\mathbf{F}$, were obtained by regressing the matrix of variable loadings $\mathbf{V}$ onto the target vectors $\mathbf{T}$, which has 26 rows and 5 columns (i.e., one for each class).

$$\mathbf{F} = \mathbf{VT}(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1} \tag{7}$$

The factor spectrum for each class is a column of $\mathbf{F}$.

Experimental Section

Reagents and Instrumentation

All chemicals were purchased from Sigma Chemical Co. (St. Louis, MO) and were used without further purification. *Bacillus anthracis* (spores and cells), *Francisella tularensis*, *Yersinia pestis*, and *Brucella melitensis* were obtained as gamma-killed freeze-dried cells from the Armed Forces Institute of Pathology (Washington, DC). These samples were prepared as 10 mg/mL water suspensions. For mass spectral analysis, 15 μL of the bacterial suspension was coinjected with 5 μL of 1 M tetramethylammonium hydroxide (TMAH) in water immediately prior to pyrolysis [11, 21]. TMAH and similar derivatization reagents have been used on bacteria to chemically lyse the biopolymers and increase the volatility of analytes, allowing analysis by gas chromatography and mass spectrometry [22, 23]. This reagent also enhances membrane permeability, as the methylation process forms products that are less polar than the underivatized species [21].

All analyses were performed using a chemical/biological mass spectrometer (CBMS) that is an air buffered quadrupole ion trap mass spectrometer fitted with an infrared pyrolyzer (Bruker-Franzen, Bremen, Germany) [21]. See Figure 1 for a schematic of the instrument. The quartz pyrolysis chamber is a 4 cm tube with an i.d. of 3 mm and an o.d. of 4 mm. Centered in this chamber is a quartz frit (3 mm thickness), upon which the sample is collected prior to pyrolysis. During thermal hydrolysis and methylation, a temperature of 450 °C was maintained for 55 s. THM products were transferred through a three meter deactivated fused silica transfer line using air as the carrier gas at a flow rate of 1.5 L/min. All thermal reactions were performed in air at atmospheric pressure. The transfer line was held at a temperature of 180 °C and the furnace base
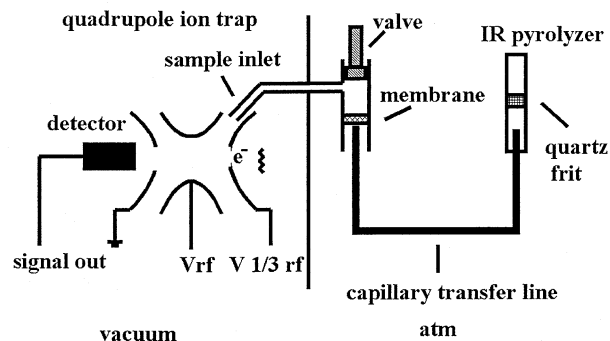
Figure 1. Schematic of the chemical/biological mass spectrometer (CBMS).

unit was held at a temperature of 200 °C. A temperature of 200 °C was used for the silicone membrane that interfaces the transfer line to the ion trap mass spectrometer. The quadrupole ion trap pressure was maintained at 5.0×10^{-5} torr. Electron ionization (70 eV) was used for all analyses. Ionization times were automatically controlled and varied from 50 μs to 15 ms to prevent space charge effects from occurring in the trap. An ion cooling time of 50 μs was used prior to scanning.

Hardware and Software

The TTFA and TCCCN programs were written in C++ and compiled using Borland C++ 5.02. The code was executed on a 450 MHz Pentium II computer with 128 MB of RAM that operated under Windows NT 4.0 SP6. All calculations used single-precision (32 bit) floating-point arithmetic. The TCCCNs trained within 2 min for both the single network models with multiple outputs and the multiple network models with single outputs. Random numbers were obtained from the uniform random number function and were then randomly shuffled [20]. The random seed values for the random number generator were 10 digit random numbers.

Target Transformation Factor Analysis

The principal components were calculated using SVD [20]. The number of principal components was calculated so that 99% of the variance in the mass spectral data set would be spanned. The number of components that spanned 99% of the variance was 26. Other numbers of components (i.e., 20 and 156) only marginally affected the factor spectra. The TTFA calculated five factors that characterized class differences. A pseudoinverse $\mathbf{T}^+$ calculated using a second SVD computation replaced $\mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}$ in eq 7. The pseudoinverse used one ppm of the summed singular values as its threshold criterion.

Evaluation of the Neural Network Models

The 214 pyrolysis mass spectra were grouped into the five classes that are given in Table 1. Each bacterial class

Table 1. THM-MS data of the five bacteria classes with species and sub-species

Class	Sample	Number of spectra
A	*Bacillus anthracis* cells	33
	Vollum	
	Zimbabwe	
	Ames	
	Sternes	
S	*Bacillus anthracis* spores	32
	Vollum	
	Zimbabwe	
	Ames	
	Sternes	
B	*Brucella melitensis*	60
	melitensis/Wild	
	melitensis/Rev-1	
	Suis	
	abortus/Wild	
	abortus/S19/vac	
F	*Francisella tularensis*	43
	Type A/Utah	
	Palaeartica	
	LVS	
Y	*Yersinia pestis*	46
	195/P India	
	La Paz	
	Nair Kenya	
	A1122 California	
	EV76	
Total		214

comprised several sub-species. In addition, a two level factorial design was used with the bacteria cultivated on different media and different culture times. Each sample (i.e., bacterial subspecies, media, and culture time) furnished four spectra collected at different times. Ten spectra were removed from this data set, because they were of poor quality. Further details may be found in Hendricker et al. [21]. The *Bacillus anthracis* spores and cells were grouped into separate classes because their mass spectra were different. The data were partitioned into training and test sets using the Latin-partition method [14]. The test sets were used to evaluate the predictability of the neural network models.

Every neural network was trained so that a model was obtained once the 10% RRMSEC criterion was satisfied. The TCCCNs converged for all cases to a relative training error below 10%. No TCCCN models were excluded because they performed poorly.

The Latin-partition method [14] builds sets of training and test set pairs so that each spectrum occurs once and only once among the partitioned test sets. The advantage of this method is that the entire set of spectra is evaluated as opposed to a subset. The method randomly draws spectra without replacement from the training set for each test set. Each class is represented in the same proportions in the training and test sets that are found in the entire data set. The effects of partitioning and the neural network training were evaluated by

using different seeds for the random number generation in the partitioning algorithm and the initial neural network models.

In this work, the effect of randomly partitioning the data into training and test sets was investigated. A Latin-partition split the data so that 75% of the spectra comprised a training set and 25% of the spectra comprised the test set. Every spectrum in the entire set of data was used once and only once among the four test sets obtained from these partitions. The prediction results of the four test sets were pooled to yield statistics that characterized the entire set of 214 spectra. The partitioning was evaluated with five replications, so that the effect of the random drawing of the data could be evaluated. This set of experiments resulted in twenty neural network models and evaluations.

Results and Discussion

Data Preprocessing

The mass range between m/z 146 and 398 was used because this range was optimal for detecting characteristic bacterial fatty acid methyl ester (FAME) peaks. This range avoids mass spectral peaks that correspond to the TMAH reagent. Before training and testing, each spectrum was normalized so that it had a unit vector length. The network models identified classes with output values that were greater than or equal to 0.5, and output values less than 0.5 indicated that a class was not present.

Principal Component Analysis

Principal component analysis was applied to the entire set of mass spectra. The scores of the 214 objects of five classes of bacterial mass spectra using the first two principal components are given in Figure 2. The principal component scores were obtained from SVD of the mean-centered data set. The spectral scores on additional components did not indicate any further clustering of the bacteria, which could be expected because the experimental factors of sub-species, culture times, and growth media were varied for each bacterial class. The
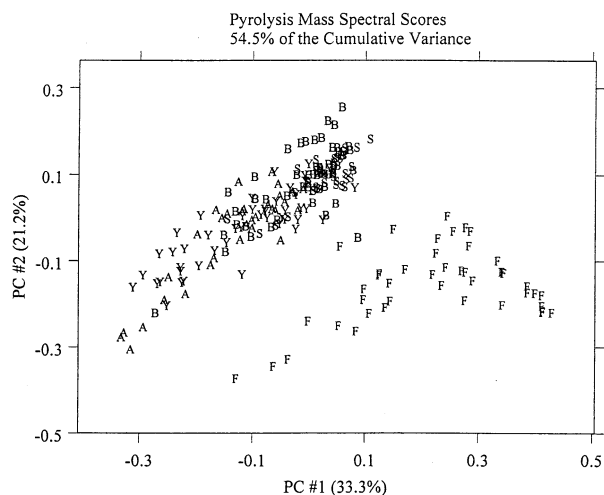

Figure 2. Scores of the THM-MS objects on the first two principal components of the THM-MS data set: *Bacillus anthracis* cells (A) and spores (S), *Brucella melitensis* (B), *Yersinia pestis* (Y), and *Francisella tularensis* (F). The percent variance spanned by each component is given with the axis label.

Francisella tularensis (F) spectra are almost linearly separable from the other classes with the exception of a single *Brucella melitensis* (B) spectrum; however, no linearly separable region is observed among the other classes for the first two principal components observation scores.

Neural Network Classification Models

The first set of experiments used Latin partitions to create four training-test set pairs. The partitioning and network training processes were run five times so that twenty prediction set results were obtained. Each neural network output unit corresponds to a target class. The prediction results were pooled for each set of four partitions. The results are presented in Table 2 for single neural network models with five outputs. The figures of merit (FOM) in Table 2 are described in the following paragraphs.

The root mean square standard error of prediction (RMSEP) is calculated by a similar formula as the

Table 2. Results from 5 sets of Latin partitions using a 75%/25% split of the training and testing data

	RMSEP[b]	RRMSEP[c]	FP[d]	FN[e]	HI[f]
Run 1	0.11	0.28	8	7	207
Run 2	0.11	0.26	6	8	206
Run 3	0.09	0.22	2	6	208
Run 4	0.14	0.35	8	14	200
Run 5	0.13	0.33	10	10	204
Average[a]	0.11 ± 0.02	0.29 ± 0.06	6.8 ± 3.8	9.0 ± 3.9	205.0 ± 3.9

[a]The average results for the prediction with multiple output TCCCN models are accompanied with a 95% confidence interval.
[b]Root mean standard error of prediction.
[c]Relative root mean square error of prediction.
[d]Number of false positive.
[e]Number of false negative.
[f]Number of correct identifications.

Table 3. Results from 5 sets of Latin partitions using a 75%/25% split of the training and testing data

	RMSEP	RRMSEP	FP	FN	HI
Run 1	0.12	0.31	13	8	206
Run 2	0.14	0.34	11	13	201
Run 3	0.10	0.26	3	8	206
Run 4	0.14	0.35	13	12	202
Run 5	0.14	0.36	13	13	201
Average[a]	0.13 ± 0.02	0.32 ± 0.05	10.6 ± 5.4	10.8 ± 3.2	203.2 ± 3.2

[a]The average results for the prediction with single output TCCCN models are accompanied with a 95% confidence interval.

RRMSEP and RRMSEC in eq 1 except the denominator is the product of the number of objects n and classes p, and the prediction results from the test sets are used instead of the estimates from the calibration sets. RMSEP is a good measure of the selectivity of the neural network model. The RMSEP will be low when the correct target class output value is 1.0 and all the other class outputs are zero.

In addition, the same data sets were used to compare prediction results with a reference method. Discriminant partial least squares (DPLS) [24] was used to compare prediction results with the TCCCN. DPLS models used the same classification rules and preprocessing as the TCCCN models. Because TCCCN models were not optimized, a reference method such as DPLS was used to verify that the TCCCN classification performance was satisfactory. The PLS program was written in C++ and implemented the orthogonal PLS-2 algorithm [25, 26]. The number of latent variables was selected to yield the minimum RMSEP for each set for the four training-test set pairs. The resulting DPLS FOMs were biased for the Latin-partitions, but yield a good lower bound for the prediction errors.

The average value for the DPLS RMSEPs for the five Latin partitions was 0.17 ± 0.04. The precision is a 95% confidence level obtained from a t-statistic. In Table 2, the TCCCN furnished a lower RMSEP of 0.11 ± 0.02, and the TCCCN prediction results were unbiased and obtained from neural network models that were not optimized but trained to a predetermined RRMSEC of 10%. A reason the TCCCN furnishes lower RMSEP values is that the sigmoidal output units can fit the binary structure of encoded classes, which is more difficult for PLS and other linear classifiers. The other biased PLS classification FOMs were statistically similar to the results reported in Table 2 and Table 3.

The neural network outputs ranged between 0 and 1. A classification criterion of 0.5 was used, so that outputs that were greater than or equal to 0.5 indicated class membership. Therefore, three informative statistics are the number of false positives (FP), false negatives (FN), and the hit index (HI). A false positive occurs when a bacterial class is misidentified as present (i.e., the network output is greater than or equal to 0.5 and the target value is 0). A false negative occurs when a bacterial class that is present is not recognized by the neural network model (i.e., the network output is less

than 0.5 and the target value is 1). The HI is a measure of correct recognition (i.e., the network output is greater than 0.5 and the target value is 1). The sum of HI and FN should equal the total number of spectra for these data sets. All the averages are accompanied by 95% confidence intervals that characterize the variations of partitioning the test sets and training the neural network models with different randomized weights.

A poorly conditioned class of bacterial samples could have a detrimental impact on the entire neural network model, thus having multiple networks with single outputs may provide better classification. Because each class has its own network, the poorly conditioned class would be isolated and not introduce errors into the other classifications.

A second evaluation was run on the same design. For this evaluation, five neural network models were used with a single output. The run numbers correspond to the same sets of training-test set pairs for the two evaluations; in this case, five neural network models for which each network model recognizes a single target class. The prediction results of this study are presented in Table 3.

The results from the two studies are interesting because they indicate that using multiple networks with single outputs provided no tangible benefits for these data. The results indicate a marginal improvement for using a single network with multiple outputs. However, a paired t-test revealed that the multiple output network model was significantly better with 95% confidence for the FP statistic and 90% confidence for the RMSEP values.

The prediction results for the five Latin-partitions of the multiple output TCCCN (i.e., reported in Table 2) were rearranged into a confusion matrix. The confusion matrix comprises rows that correspond to the target classes and columns that correspond to the TCCCN output for the target class [27]. The number of correct classifications for each class is located along the diagonal of the matrix. The numbers of misclassifications are located in the off-diagonal elements. When a spectrum is misclassified, the mistaken bacterial class is indicated by the column of the off-diagonal result. If two classes of bacteria are confused by the model, then large values will occur on the off-diagonal elements indicated by the rows of the true bacterial classes and columns of the mistaken classes.

Table 4. Confusion matrix from the 5 sets of Latin partitions using a 75%/25% split of the training and testing data[a]

True	A est.	S est.	B est.	F est.	Y est.	Null est.
A	32 ± 0.9	0.6 ± 1.1	0	0	0.2 ± 0.6	0.2 ± 0.6
S	1.4 ± 0.7	29.6 ± 1.4	0	0	0.8 ± 1.4	0.2 ± 0.6
B	0	0	58.4 ± 1.9	0.6 ± 0.7	1 ± 1.8	0.2 ± 0.6
F	0.2 ± 0.6	0.2 ± 0.6	0	43 ± 0.0	0	0
Y	0.8 ± 1.0	0	0.4 ± 0.7	0	43.4 ± 1.9	1.4 ± 1.4

[a]The average results for the prediction with multiple output TCCCN models are accompanied with a 95% confidence interval. Each row represents the true class and each column represents the average output classification across the five replicated partitions. Null estimate indicates that no classification was made for an object.

Because outputs from the TCCCN models greater than 0.5 are used to classify bacteria as opposed to using the largest output from the neural network, a single spectrum may result in several misclassifications. In addition, a spectrum may not be classified into any of the target classes when all of the TCCCN outputs are less than 0.5. Thus, a column labeled null estimate indicates the number of times no class assignment is made for each class. For poor quality spectra, a null estimate is preferable to wrongly classifying the spectrum.

The confusion matrix presented in Table 4 contains the average prediction results for the five partitions. A value of 0.2 would indicate that a misclassification value occurred only once during the five evaluations and a value of 0.6 would indicate that a misclassification occurred three times for the five evaluations.

In Figure 2, one spectrum for *Brucella melitensis* (B) at approximate coordinates of 0.1, −0.1, is located close to the *Francisella tularensis* (F) cluster, which makes it an obvious candidate for misclassification. This point was misidentified in all five of the PLS Latin-partitions; however, it was correctly classified by a single run (i.e., Latin-partition) with the TCCCN, misclassified by three runs, and in one run no classification was made for this spectrum. These results appear in Table 4 for which the three misclassifications for this spectrum were for *Francisella tularensis* (F) and one occurred as a null estimate (i.e., no classification).

Assessment of Important Mass Spectral Features

A method for identifying the important variables from TCCCN models has been devised using sensitivities [17]. Characteristic features for each class are variables that have large sensitivities. Five TCCCN models were built by using the entire set of THM-MS data, training the TCCCN to a 10% RRMSEC and using different random starting conditions. The sensitivity for each class was scaled by its largest sensitivity value (i.e., normalized to a maximum intensity of unity) to furnish relative sensitivities. The relative sensitivities for each class from the different TCCCN models were averaged and plotted with respect to *m/z* values. The larger absolute sensitivities are given in Tables 5, 6, 7, 8, and 9. In many instances, the sensitivities would be similar but differ in scale, which yielded the larger confidence intervals. This approach was used to remove absolute variations among scale from the replicate sensitivity measurements. The size of the sensitivities indicates the relative importance of mass spectral peaks for the TCCCN classification. The positive features indicate peaks that are characteristic for a class; the negative features indicate peaks that are characteristic of the other classes in the data set. A model linearity test was applied to the TCCCN classification models as described in reference [17]. For all classes this test for linearity indicated that TCCCN models were simple and linear.

There are several steps to validating the neural network model. First, the TTFA spectrum should be compared to the sensitivity spectrum. The two should have some peaks in common and peaks that point in similar directions. Bad neural network models that overfit the training mass spectra set will give sensitivity spectra that resemble white noise or are very noisy [17]. The next step is to examine the *m/z* of the larger peaks in the sensitivity spectrum for correspondence to

Table 5. Largest average sensitivities with 95% confidence intervals for *Bacillus anthracis* cell class (A) and single TCCCN multiple output models

m/z	Absolute sensitivity	Relative sensitivity	TTFA loading	Probable source[a]
168	(2.05 ± 1.88) × 10^{-5}	100 ± 0%	44%	Protein
188	(2.02 ± 1.95) × 10^{-5}	94 ± 16%	85%	Protein
162	(4.10 ± 6.69) × 10^{-6}	47 ± 54%	29%	
163	(3.53 ± 3.24) × 10^{-7}	20 ± 43%	100%	
193	(1.85 ± 3.38) × 10^{-7}	0.5 ± 9%	62%	
194	(5.99 ± 9.24) × 10^{-8}	0.9 ± 1.8%	29%	Protein

[a]The source for each peak is postulated on published information.

Table 6. Largest average sensitivities with 95% confidence intervals for *Bacillus anthracis* spores class (S) and single TCCCN multiple output models and the corresponding relative TTFA loading

m/z	Absolute sensitivity	Relative sensitivity	TTFA loading	Probable source[a]
196	$(3.98 \pm 3.67) \times 10^{-5}$	$100 \pm 0\%$	100%	Dimethyl dipicolinic acid
215	$(3.98 \pm 3.67) \times 10^{-5}$	$100 \pm 0\%$	37%	Protein
166	$(3.98 \pm 3.67) \times 10^{-5}$	$100 \pm 0\%$	40%	
241	$(3.98 \pm 3.67) \times 10^{-5}$	$100 \pm 0\%$	10%	Lipid or protein
216	$(3.98 \pm 3.67) \times 10^{-5}$	$100 \pm 0\%$	18%	Protein
202	$(3.97 \pm 3.66) \times 10^{-5}$	$100 \pm 0\%$	22%	Protein
163	$(3.97 \pm 3.69) \times 10^{-5}$	$94 \pm 9\%$	31%	
164	$(3.96 \pm 3.65) \times 10^{-5}$	$91 \pm 2\%$	11%	
188	$(3.93 \pm 3.62) \times 10^{-5}$	$99 \pm 1\%$	−74%	Protein
174	$(3.93 \pm 3.61) \times 10^{-5}$	$96 \pm 5\%$	5%	Protein or lipid

[a]The source for each peak is postulated on published information.

Table 7. Largest average sensitivities with 95% confidence intervals for *Brucella melitensis* (B) class and single TCCCN multiple output models and the corresponding relative TTFA loadings

m/z	Absolute sensitivity	Relative sensitivity	TTFA loading	Probable source[a]
235	$(3.38 \pm 3.11) \times 10^{-5}$	$100 \pm 0\%$	100%	Protein
249	$(3.38 \pm 3.11) \times 10^{-5}$	$100 \pm 0\%$	52%	Protein
161	$(3.38 \pm 3.11) \times 10^{-5}$	$100 \pm 0\%$	22%	Protein
175	$(3.37 \pm 3.11) \times 10^{-5}$	$100 \pm 0\%$	23%	Protein
179	$(3.37 \pm 3.11) \times 10^{-5}$	$100 \pm 0\%$	25%	
207	$(3.37 \pm 3.10) \times 10^{-5}$	$100 \pm 0\%$	12%	Protein
148	$(3.37 \pm 3.10) \times 10^{-5}$	$100 \pm 0\%$	17%	
189	$(3.37 \pm 3.10) \times 10^{-5}$	$100 \pm 0\%$	28%	Protein
149	$(3.36 \pm 3.10) \times 10^{-5}$	$100 \pm 0\%$	25%	
264	$(3.35 \pm 3.09) \times 10^{-5}$	$99 \pm 1\%$	28%	Lipid

[a]The source for each peak is postulated on published information.

Table 8. Largest average sensitivities with 95% confidence intervals for *Francisella tularensis* (F) class and single TCCCN multiple output models and the corresponding relative TTFA loadings

m/z	Absolute sensitivity	Relative sensitivity	TTFA loading	Probable source[a]
199	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	100%	Lipid
185	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	59%	Lipid
255	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	59%	Lipid
213	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	30%	Lipid
227	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	28%	Lipid
241	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	22%	Lipid
297	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	39%	Lipid
269	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	33%	Lipid
157	$(3.25 \pm 3.00) \times 10^{-5}$	$100 \pm 0\%$	34%	Lipid
298	$(3.25 \pm 2.99) \times 10^{-5}$	$100 \pm 0\%$	30%	Lipid

[a]The source for each peak is postulated on published information.

Table 9. Largest average sensitivities with 95% confidence intervals for *Yersinia pestis* class (Y) and single TCCCN multiple output models and the corresponding relative TTFA loadings

m/z	Absolute sensitivity	Relative sensitivity	TTFA loading	Probable source[a]
179	$(0.62 \pm 1.33) \times 10^{-5}$	$20 \pm 56\%$	20%	
148	$(0.60 \pm 1.29) \times 10^{-5}$	$19 \pm 54\%$	−9%	Nucleic acid
233	$(0.45 \pm 1.00) \times 10^{-6}$	$2 \pm 4\%$	46%	Protein
160	$(4.22 \pm 6.27) \times 10^{-7}$	$2 \pm 5\%$	34%	Protein
178	$(1.79 \pm 2.27) \times 10^{-7}$	$67 \pm .58\%$	100%	Nucleic acid
214	$(4.68 \pm 6.96) \times 10^{-8}$	$39 \pm 58\%$	31%	Protein
158	$(3.84 \pm 5.17) \times 10^{-8}$	$31 \pm 51\%$	66%	Protein

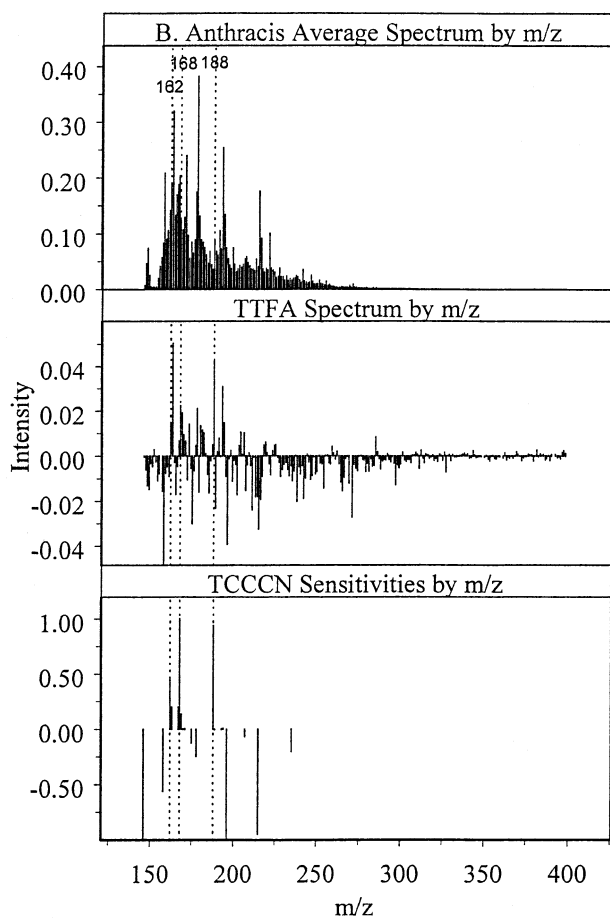[a]The source for each peak is postulated on published information.

Figure 3. The average THM-MS spectrum for *B. anthracis* (A), the TTFA spectrum, and the average sensitivity spectrum obtained from the TCCCN model.

Figure 4. The average THM-MS spectrum for *B. anthracis* spores (S), the TTFA spectrum, and the average sensitivity spectrum obtained from the TCCCN model.

known biomarker peaks. For example, peaks corresponding to pump oil appearing in the sensitivity spectrum would be undesirable. Lastly, the confidence intervals of the larger peaks in the sensitivity spectra indicate the stability and the reproducibility of the TCCCN models. They also indicate whether the intensity of the peak is important for the classification. Negative peaks in both the TTFA and TCCCN spectra are characteristic for the other bacterial classes.

The class average spectrum, the TTFA spectrum, and the TCCCN normalized averaged sensitivity spectrum for the *Bacillus anthracis* vegetative spectra is given in Figure 3. The low mass range of *m/z* 150–250 contains the characteristic peaks in both the TTFA loadings and in the neural network model sensitivities. The difference between the significant peaks in sensitivity and the TTFA spectra can be seen in Table 5. This class and the *Yersinia pestis* class were two classes that exhibited a large degree of variability among the relative sensitivities. They also were the two classes with a few sensitivities dominating the classification model. The variability in sensitivity is caused by model ambiguity because only a few peaks in the spectrum are needed for classification. The neural network models used
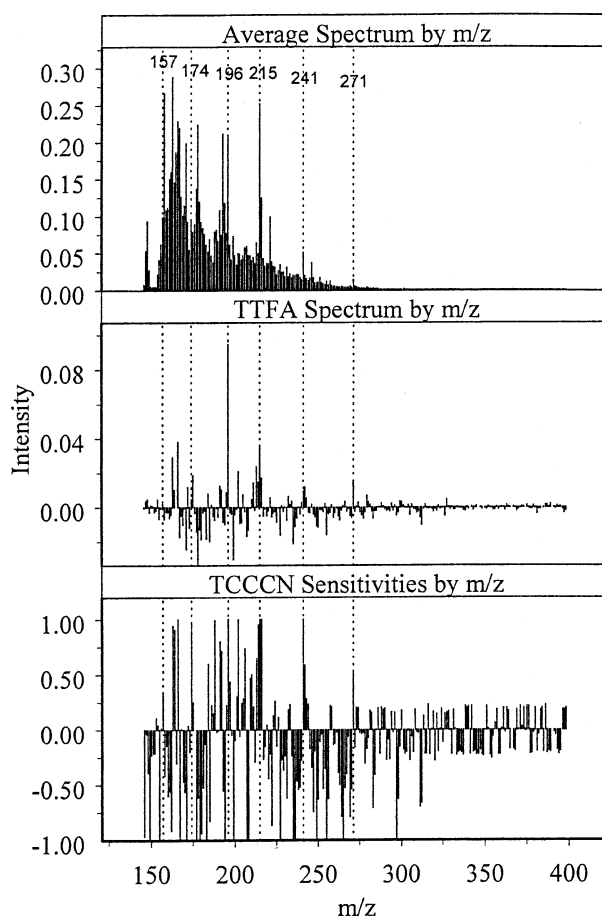
different peaks for replicate models to classify these two sets of bacteria (A and Y).

The *Bacillus anthracis* classes (A and S) gave the sensitivity spectra with the largest average values (see Tables 5 and 6). For the *Bacillus anthracis* spores, the peak at *m/z* 196 in Table 6 and Figure 4 corresponds to the $(M + H)^+$ ion of the dimethylated derivative of dipicolinic acid. Dipicolinic acid is a well-recognized and common component of sporulated bacilli. The TTFA variable loadings in Figure 4 contain small peaks at high mass (*m/z* 300–400).

Good agreement was obtained for the *Brucella melitensis* sensitivity spectrum and the TTFA loadings in Figure 5. The largest averaged sensitivities are listed in Table 7. Sensitivities that all have the same intensity indicate that the corresponding intensities of the mass spectral peaks were unused by the TCCCN classification model. Because many of the peaks in the sensitivity spectrum are at 100% indicate that peak intensity, the corresponding spectral peaks in the input spectrum were used qualitatively for discrimination (i.e., only the existence and not the intensity of the peak was used to classify the spectrum). Most of these peaks correspond to peptide ions.
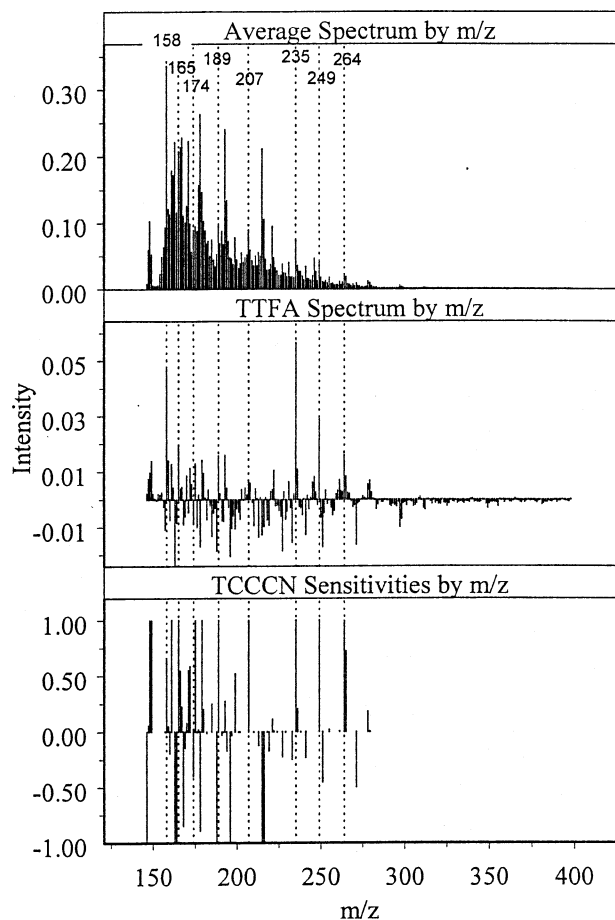
Figure 5. The average THM-MS spectrum for *B. melitensis* (B), the TTFA spectrum, and the average sensitivity spectrum obtained from the TCCCN model.
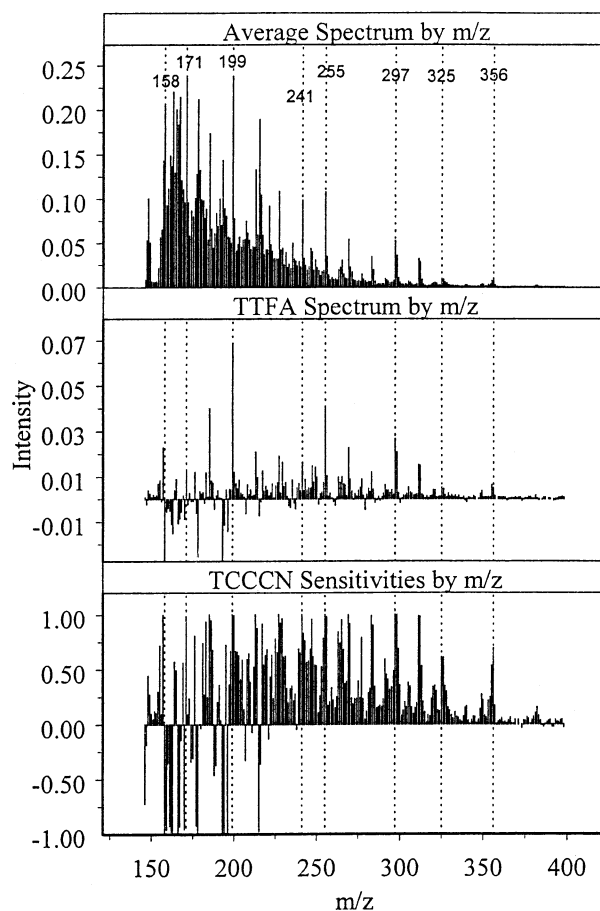

Figure 6. The average THM-MS spectrum for *F. tularensis* (F), the TTFA spectrum, and the average sensitivity spectrum obtained from the TCCCN model.

For *Francisella tularensis* both the TTFA and sensitivity spectra indicate that the m/z range from 200–350 contains much of the important information classification (see Figure 6). Table 8 gives a comparison between the average sensitivities, average relative sensitivities, and TTFA results. The characteristic sensitivities correspond to saturated fatty acid methyl esters and carbomethoxy fragmentation series.

The sensitivities had very good reproducibility for every class except for the *Yersinia pestis* class (see Table 9 and Figure 7). In the TCCCN models, only a few sensitivities values were large. This class exhibited a large degree of variability among the TCCCN models obtained from different random conditions. For the first TCCCN model, m/z 160 and 178 were the two larger sensitivities. For the second TCCCN model, m/z 179 and 148 were the two larger sensitivities. For the third through fifth TCCCN models, excellent repeatability was achieved with the largest sensitivities at m/z 178 and 215. The results reported in Table 9 and Figure 6 were average sensitivities of these five models. The peak at m/z 178 is a DNA base peak and the peak at m/z 215 is a protein oxidation product. There was an intense negative sensitivity peak at m/z 162 that dominated in

magnitude four of the five models that was used for distinguishing *Yersinia pestis* (Y) spectra from the *Bacillus anthracis* (A) spectra. For these two classes, the mass spectral peaks over m/z 200 were not useful for classification. The lack of discerning peaks in the spectra may explain the larger variability in the replicated sensitivity spectra for these two classes.

Conclusions

Target transform factor analysis, a linear multivariate tool for exploratory analysis, was demonstrated for validating neural network models and visualizing features in complex multivariate data sets. The neural network model should no longer be considered a black-box classifier. Instead, using sensitivity analysis allows the relationships among the input variables and the output properties to be studied. Prediction results and the test of linearity demonstrated that no benefits were obtained from multiple network single output models. For some classes of bacteria different mass ranges were used for classification by the TCCCN models. In other cases, the magnitudes of the mass spectral peaks were not utilized by the neural network models. Discarding
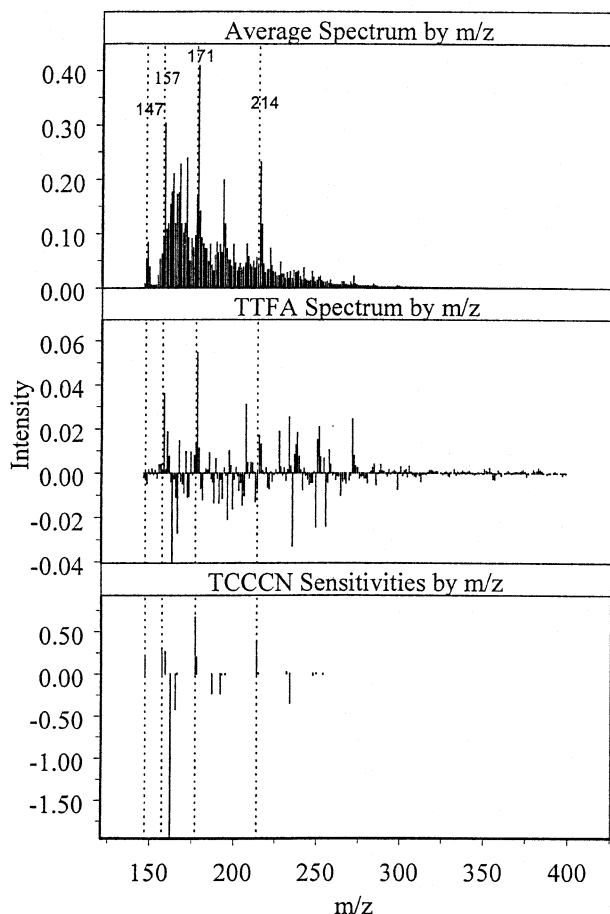
Figure 7. The average THM-MS spectrum for *Y. pestis* (Y), the TTFA spectrum, and the average sensitivity spectrum obtained from the TCCCN model.

Single networks comprised of multiple outputs furnished simpler classification models that were less susceptible to variations introduced in partitioning the calibration and test sets. Both the multiple network models with single outputs and single network models took similar amounts of time (i.e., less than 2 min on a personal computer) to train for these data sets. The sensitivities of the TCCCN models were similar for both the single output and multiple output networks and agreed with the variable loadings obtained from TTFA using the class designations as targets. Furthermore, the sensitivity spectra indicate that the neural network models are using mass spectral features for classification that have been identified as biomarkers in previous studies. The TCCCN provides a robust, reliable, and rapid tool for solving complex classification problems. Sensitivity spectra provide useful diagnostics for assessing neural network models and the causal relationships between the input data and the target classes.

Acknowledgments

magnitude information is an attribute of nonlinear classifiers. Furthermore, characteristic features were recognized as originating from protein, lipid, or nucleic acid sources and corresponded to other known biomarkers.

The neural network models recognized the target bacteria from their THM-MS spectra with 96 ± 2% accuracy and had a false alarm rate of 3 ± 2%. The measures of precision are 95% confidence intervals. These results were obtained without the time-consuming process of optimizing the neural networks, but instead training to a reasonable classification threshold of a 10% RRMSEC. In addition, the TCCCN models yielded significantly lower prediction errors than PLS models that were optimized to yield the lowest possible prediction errors.

For visualizing and discovering characteristic features in intricate multivariate data sets, TTFA was demonstrated as a useful tool. The sensitivity models corresponded well with the TTFA spectra for the bacteria classes. A procedure for calculating relative sensitivities was developed to control a main source of variation of scale that occurs among replicate TCCCN models.

References

1. Goodacre, R.; Trew, S.; Wrigleyjones, C.; Neal, M. J.; Maddock, J.; Ottley, T. W.; Porter, N.; Kell, D. B. Rapid Screening for Metabolite Overproduction in Fermentor Broths Using Pyrolysis Mass-Spectrometry with Multivariate Calibration and Artificial Neural Networks. *Biotechnol. Bioeng.* **1994,** *44,* 1205–1216.

2. Goodacre, R.; Neal, M. J.; Kell, D. B. Rapid and Quantitative Analysis of the Pyrolysis Mass Spectra of Complex Binary and Tertiary Mixtures Using Multivariate Calibration and Artificial Neural Networks. *Anal. Chem.* **1994,** *66,* 1070–1085.

3. Goodacre, R.; Kell, D. B. Correction of Mass Spectral Drift Using Artificial Neural Networks. *Anal. Chem.* **1996,** *68,* 271–280.

4. Goodare, R.; Rooney, P. J.; Kell, D. B. Discrimination Between Methicillin-Resistant and Methicillin-Susceptible *Staphylococcus aureus* Using Pyrolysis Mass Spectrometry and Artificial Neural Networks. *J. Antimicrob. Chemother.* **1998,** *41,* 27–34.

5. Goodare, R.; Timmins, E. M.; Burton, R.; Kaderbhai, N.; Woodward, A. M.; Kell, D. B. Rapid Identification of Urinary Tract Infection Bacteria Using Hyperspectral Whole-Organism Fingerprinting and Artificial Neural Networks. *Microbiol. UK* **1998,** *144,* 1157–1170.

6. Sission, P. R.; Freeman, R.; Law, D.; Ward, A. C.; Lightfoot, N. F. Rapid Detection of Verocytotoxin Production Status in *Escherichia coli* by Artificial Neural-Network Analysis of Pyrolysis Mass-Spectra. *J. Anal. Appl. Pyrolysis* **1995,** *32,* 179–185.

7. Freeman, R.; Sisson, P. R.; Ward, A. C. Resolution of Batch Variation in Pyrolysis Mass-Spectrometry of Bacteria by the Use of Artificial Neural-Network Analysis. *Antonie van Leeuwenhoek Int. J. Gen. Molec. Microbiol.* **1995,** *68,* 253–260.

8. Kenyon, R. G. W.; Ferguson, E. V.; Ward, A. C. Application of Neural Networks to the Analysis of Pyrolysis Mass Spectra.

Zentralblatt Für Balterologie-Int. J. Med. Microbiol. Viro. Parasito. Infect. Diseases **1997,** *285,* 267–277.

9. Beverly, M. B.; Basile, F.; Voorhees, K. J. Fatty Acid Analysis of Beer Spoiling Microorganisms Using Pyrolysis Mass Spectrometry. *J. Am. Soc. Brew. Chem.* **1997,** *55,* 79–82.

10. Basile, F.; Beverly, M. B.; Voorhees, K. J.; Hadfield, T. L. Pathogenic Bacteria: Their Detection and Differentiation by Rapid Lipid Profiling with Pyrolysis Mass Spectrometry. *Trac-Trends Anal. Chem.* **1998,** *17,* 95–109.

11. Basile, F.; Beverly, M. B.; Abbas-Hawks, C.; Mowry, C. D.; Voorhees, K. J.; Hadfield, T. L. Direct Mass Spectrometric Analysis of in situ Thermally Hydrolyzed and Methylated Lipids from Whole Bacterial Bells. *Anal. Chem.* **1998,** *70,* 1555–1562.

12. Harrington, P. B. Temperature-Constrained Cascade Correlation Networks. *Anal. Chem.* **1998,** *70,* 1297–1306.

13. Cai, C.; Harrington, P. B. Prediction of Substructure and Toxicity of Pesticides with Temperature Constrained-Cascade Correlation Network from Low-Resolution Mass Spectra. *Anal. Chem.* **1999,** *71,* 4134–4141.

14. Wan, C.; Harrington, P. B. Screening GC-MS Data for Carbamate Pesticides with Temperature Constrained-Cascade Correlation Networks. *Anal. Chim. Acta* **2000,** *408,* 1–12.

15. Fahlman, S. E.; Lebiere, C. The Cascade-Correlation Learning ArchitectureCMU-CS-90-100. Carnegie Mellon University, 1991, 1–13.

16. Harrington, P. B. Temperature Constrained Backpropagation Neural Networks. *Anal. Chem.* **1994,** *66,* 802–807.

17. Harrington, P. B.; Wan, C.; Urbas, A. Generalized Sensitivity Analysis of Neural Network Models. *Anal. Chem.* **2000,** *72,* 5004–5013.

18. Malinowski, E. R.; *Factor Analysis in Chemistry 2nd ed.* John Wiley & Sons, Inc: New York, 1991, pp. 22–31.

19. Hopke, P. K. Target Transformation Factor Analysis. *Chemomet. Intel. Lab. Syst.* **1989,** *6,* 7–19.

20. Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C, 2nd ed.* Cambridge University Press: Cambridge, 1994 (SVD 59–70; Random Number 146–159).

21. Hendricker, A. D.; Abbas-Hawks, C.; Basile, F.; Voorhees, K. J.; Hadfield, T. L. Rapid Chemotaxonomy of Pathogenic Bacteria Using in situ Thermal Hydrolysis and Methylation as a Sample Preparation Step Coupled with a Field-Portable Membrane-Inlet Quadrupole Ion Trap Mass Spectrometer. *Int. J. Mass Spectrom.* **1999,** *191,* 331–342.

22. DeLuca, S.; Sarver, E. W.; Harrington, P. B.; Voorhees, K. J. Direct Analysis of Bacterial Fatty Acids by Curie-Point Pyrolysis Tandem Mass Spectrometry. *Anal. Chem.* **1990,** *62,* 1465–1472.

23. Kossa, W. C.; MacGee, J.; Ramachandran, S.; Webber, A. J. Pyrolytic Methylation/Gas Chromatography: A Short Review. *J. Chromatog. Sci.* **1979,** *17,* 177–187.

24. Alsberg, B. K.; Goodacre, R.; Rowland, J. J.; Kell, D. B. Classification of Pyrolysis Mass Spectra by Fuzzy Multivariate Rule Induction-Comparison with Regression, K-Nearest Neighbour, Neural and Decision-Tree Methods. *Anal. Chim. Acta* **1997,** *348,* 389–407.

25. Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986,** *185,* 1–17.

26. Martens, H.; Næs, T. *Multivariate Calibration.* John Wiley: Chichester, 1989, pp 146–158.

27. Masters, T. *Practical Neural Network Recipes in C++; 1st ed.* Academic Press: Boston, 1993, pp 348–350.

28. Wan, C. Ohio University. Ph.D., 1999, pp 112–135.