TECHNICAL NOTE:

DIFFERENTIATION OF FOOD PATHOGENS USING FTIR AND ARTIFICIAL NEURAL NETWORKS

M. J. Gupta, J. M. Irudayaraj, C. Debroy, Z. Schmilovitch, A. Mizrach

ABSTRACT. *FTIR absorbance spectra in conjunction with artificial neural networks (ANNs) were used to differentiate selected microorganisms at the generic and serogroup levels. The ANN consisted of three layers with 595 input nodes, 50 nodes at the hidden layer, and 5 output nodes (one for each microorganism or strain). Ten replications of each experiment were conducted, and 70% of the data was used for training and 30% for validation of the network. Results indicated that differentiation could be achieved at an accuracy of 80% to 100% at the generic level and 90% to 100% at the serogroup level at 103 CFU/mL concentration.*

Keywords. ANN, Differentiation, Food pathogens, FTIR spectroscopy.

Health problems due to food pathogens have been well documented (Mead et al., 1999; Crutchfield et al., 2000). Rapid monitoring of food pathogens and their identification is of great importance to food scientists, industries, as well as federal agencies. Fourier-transform infrared spectroscopy (FTIR) has been used for identification and characterization of bacteria (Heim et al., 1991a, 1991b; Naumann et al., 1991a, 1995; Curk et al., 1994; Goodacre et al., 1996; Udelhoven et al., 2000; Yang and Irudayaraj, 2003). Artificial neural networks (ANNs) have been suggested as an alternative classification method compared to the traditional multivariate statistical approach to process FTIR data due to their ability to handle nonlinear data and their tolerance of modest levels of noise or variability in the input data(Wallace et al., 2000). The ability of ANNs to learn the trends in the data and apply this knowledge to interpret new information is an attractive tool for classifying microorganism fingerprints for identification and classification purposes (Specht, 1990; Freeman et al., 1994; Chun et al., 1993; Goodacre et al., 1994, 1998; Argov et al., 2002; Kirschner et al., 1999).

The aim of this study was to study the ability of FTIR spectroscopy to differentiate a select set of foodborne pathogens at the species, strain, and various concentration levels by their absorbance spectra in the fingerprint region using ANNs.

Article was submitted for review in March 2004; approved for publication by Food & Process Engineering Institute Division of ASAE in July 2005.

The authors are **Mathala J. Gupta,** Visiting Research Scholar, Department of Agricultural and Biological Engineering, The Pennsylvania State University, University Park, Pennsylvania; **Joseph M. Irudayaraj, ASABE Member Engineer,** Associate Professor, Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana; **Chitrita Debroy,** Research Professor, The Gastroenteric Disease Center (GDC), The Pennsylvania State University, University Park, Pennsylvania; and **Ze'ev Schmilovitch, ASABE Member Engineer,** Professor, and **Amos Mizrach, ASABE Member,** Professor, Institute of Agricultural Engineering, Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel. **Corresponding author:** Joseph M. Irudayaraj, Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907; phone: 765-494-1162; fax: 765-496-1115; e-mail: josephi@psu.edu.

MATERIALS AND METHODS

BACTERIAL SAMPLE PREPARATIONS

Five pathogenic bacteria (*Enterococcus faecium, Salmonella enteritidis, Bacillus cereus, Yersinia enterocolitis*, and *Shigella*) and five *E. coli* serogroups (O103, O55, O121, O30, and O26) were obtained from the Gastroenteric Disease Center (GDC) at the Pennsylvania State University for this study. Each bacterium was cultured in 100 mL broth medium (5 g yeast extract, 8 g tryptone, and 5 g NaCl in 500 mL distilled water) at 35°C and shaken at 100 rpm for 24 h (*Enterococcus faecium, Yersinia enterocolitis, Shigella boydii*, and *E. coli*), 18 h (*Salmonella enteric* subtype Enteritidis), or 36 h (*Bacillus cereus*) to reach a final concentration of 10^9 CFU/mL.

The original samples were suspended in sterile phosphate buffer saline (PBS) and serially diluted six times (10^{-1} through 10^{-6}) such that seven different concentration levels (10^3 through 10^9 CFU/mL) could be achieved for each bacterium. All the samples (100 µL) were then plated on agar to determine the plate counts. The samples were subjected to FTIR spectroscopy measurements.

FTIR MEASUREMENTS

A BioRad Excalibur FTS 3000 spectrometer with an ATR accessory was used. FTIR absorbance spectra were collected in the spectral region between 600 and 4000 cm^{-1} at a resolution of 2 cm^{-1}. Each experiment was repeated ten times at each of the seven concentrations for the five different pathogens to give a total of 350 spectra. The same procedure was repeated for the five different *E. coli* serogroups, and 350 spectra were collected.

ARTIFICIAL NEURAL NETWORK

All ANN analyses were carried out with the Neuroshell 2 Release 4.0 package (Ward Systems Group, Inc., Frederick, Md.) using the probabilistic neural networks (PNNs) algo-

 ISSN 0001–2351

rithm, which is a model for supervised classification based on multivariate probability estimation (Specht, 1990).

The FTIR spectra were smoothed and differentiated; seven of the ten spectra from each experiment were used for training, and the remaining three were used for validating the trained network. The FTIR spectra were paired with a corresponding output using a binary encoded structure: *Enterococcus faecium* coded as 10000, *Salmonella enterica* subsp. Enteritidis as 01000, *Bacillus cereus* as 00100, *Yersinia enterocolitis* as 00010, and *Shigella boydii* as 00001. The above five training pairs collectively constituted the training set. For the *E. coli* serogroups, a similar binary coding was followed, i.e., O103 was represented as 10000, O55 as 01000, O121 as 00100, O30 as 00010, and O26 as 00001.

Initially, the absorbance data in the entire range (600 to 4000 cm^{-1}), with a total of 1764 points (or variables), were considered as input to the input layer of the neural network (NN). For such an input pattern, the training time required was between 30 and 90 min and the classification accuracy was poor, i.e., for the validation set the classification accuracy ranged from 0% to 30%, although the overall accuracy ranged from 60% to 70% (tables 1 and 2), possibly due to the presence of a high level of noise in the spectra due to interference from water and carbon dioxide. Hence, the absorbance spectra in the region 600 to 1750 cm^{-1} (Naumann, 1991b) were used as input, and only 595 points (input neurons) were implemented. The chosen region was considered sufficient because it contained the typical fingerprint (600 to 900 cm^{-1}) characterized by a combination of weak absorptions due to the aromatic ring vibrations of the phenylalanine, tyrosine, tryptophan, and several nucleotides, the polysaccharide region (900 to 1200 cm^{-1}) due to symmetric stretching vibrations and peaks due to C-O-C and C-O-P, the mixed region (1200 to 1500 cm^{-1}) due to the bending modes of lipids and proteins, and the intense amide I and amide II region due to the presence of α and β structures of the cellular proteins (Naumann, 2000).

The values applied to the input and output nodes of the 595-50-5 network were normalized, before training commenced, using the logistic function between 0 and 1 according to the built-in formula in Neuroshell 2 Release 4.0 (Ward Systems Group, Inc., Frederick, Md.). With 595 input neurons, the network was allowed to run until a minimum level of error in the test set was achieved and remained as such for 20 generations. The trained ANN was then investigated with the remaining 30% of the spectra not used for training.

Results and Discussion

Differentiation of Bacteria

To determine the optimum sample size, the training set was gradually increased from 10% to 70% of the total sample size. Figure 1 shows the effect of training set size on the classification accuracy of bacterial species. The best classification resulted with a sample size of 60% or more in the training set. Maximum error in classification was obtained for *Enterococcus faecium* and *Yersinia enterocolitis* when 10% of the training set size was used. In the same way, the error decreased as the sample size increased, and maximum prediction accuracy was achieved when 60% or more of the data was used for training. With a margin of clearance, 70% of the data was chosen for training, conforming to the norm that two-thirds of the data be used for training or calibration and one-third for validation or testing. Training of the network with a range of input data also indicated that a vast difference between the replicate sample spectra was observed, possibly due to the variability in the sample, the age of the culture, and interference due to water and carbon dioxide. This problem could be reduced significantly by increasing the number of replications in the experiments and by using as many spectra as possible for training.

Training was stopped when 20 test or calibration generations had passed and the average minimum error in classification of the patterns remained unchanged. The actual time taken to train the network was only 5 to 6 min, a significant reduction from 30 to 90 min when using the entire spectral range (600 to 4000 cm^{-1}). The trained neural network was then investigated using the data from the training set as well as the validation set from all bacteria.

Table 1. Probabilistic neural network (PNN) of bacterial species at a concentration of 103 CFU/mL in buffer.

	Input FTIR Spectra in 600 to 4000 cm^{-1} Region				Input FTIR Spectra in 600 to 1750 cm^{-1} Region			
Bacteria	Right Assignment No.	Wrong Assignment No.	Not Classified	Classified Correct (%)	Right Assignment No.	Wrong Assignment No.	Not Classified	Classified Correct (%)
Enterococcus faecium	7	0	3	70	10	0	0	100
Salmonella enteritidis	7	0	3	70	10	0	0	100
Bacillus cereus	6	0	4	60	10	0	0	100
Yersinia enterocol'tis	7	0	3	70	10	0	0	100
Shigella boydii	7	0	3	70	10	0	0	100

Table 2. Probabilistic neural network (PNN) of *E. coli* strains at a concentration of 109 CFU/mL in buffer.

	Input FTIR Spectra in 600 to 4000 cm^{-1} Region				Input FTIR Spectra in 600 to 1750 cm^{-1} Region			
E-coli Serogroup	Right Assignment No.	Wrong Assignment No.	Not Classified	Classified Correct (%)	Right Assignment No.	Wrong Assignment No.	Not Classified	Classified Correct (%)
O103	7	0	2	70	10	0	0	100
O55	7	0	3	70	10	0	0	100
O121	6	0	4	60	10	0	0	100
O30	7	0	3	70	10	0	0	100
O26	10	0	0	100	10	0	0	100

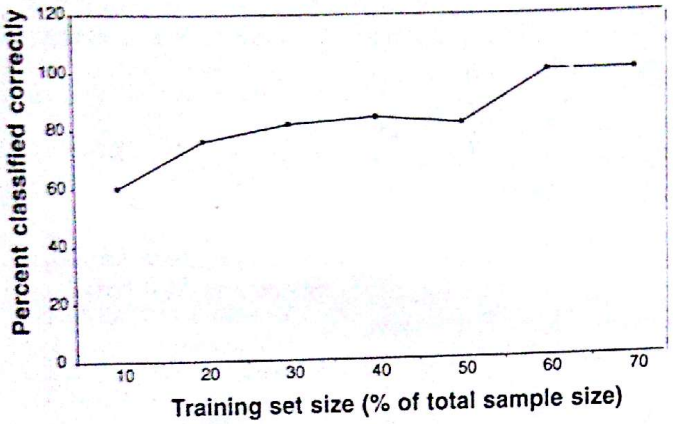

Figure 1. Effect of training set on classification of bacterial species.

Figure 2 presents the classification results of the bacterial species by ANN. Results summarized in tables 1 and 2 and figure 2 indicate that a 100% correct classification was obtained.

The level of correct species classification was between 80% and 100%. Only 10% of *Bacillus cereus* could not be classified at the 10^7 CFU/mL concentration, and 10% to 20% of *Yersinia enterocolitis* was misclassified in the concentration range 10^3 through 10^6 CFU/mL. The error in classification could be attributed to the presence of outlier patterns in the replicate spectra. Since there was no relation between the size of error and the dilution rate, it can be stated that the ANN could successfully classify the tested bacterial species at a sensitivity of 10^3 CFU/mL.

Differentiation of Serogroups

Classification of the selected *E. coli* serogroups was accomplished using an analysis similar to that described above. It was observed that 20% of the training set was sufficient to classify all four *E. coli* serogroups except the O30 strain, which was found to have a distinctly different pattern among its replicates and did not allow a correct classification for up to 50% of the training data set size. This problem was eliminated when 60% of the data set was used for training. Analysis at both the generic and serogroup levels indicated that a training set size of 60% or more of the sample was suitable. A training set size of 70% of the sample was selected to comply with the norm stated above.

Correct classification at the serogroup level was accomplished with an accuracy of 90% to 100%. Three serogroups were either wrongly classified or misclassified: O103 at the

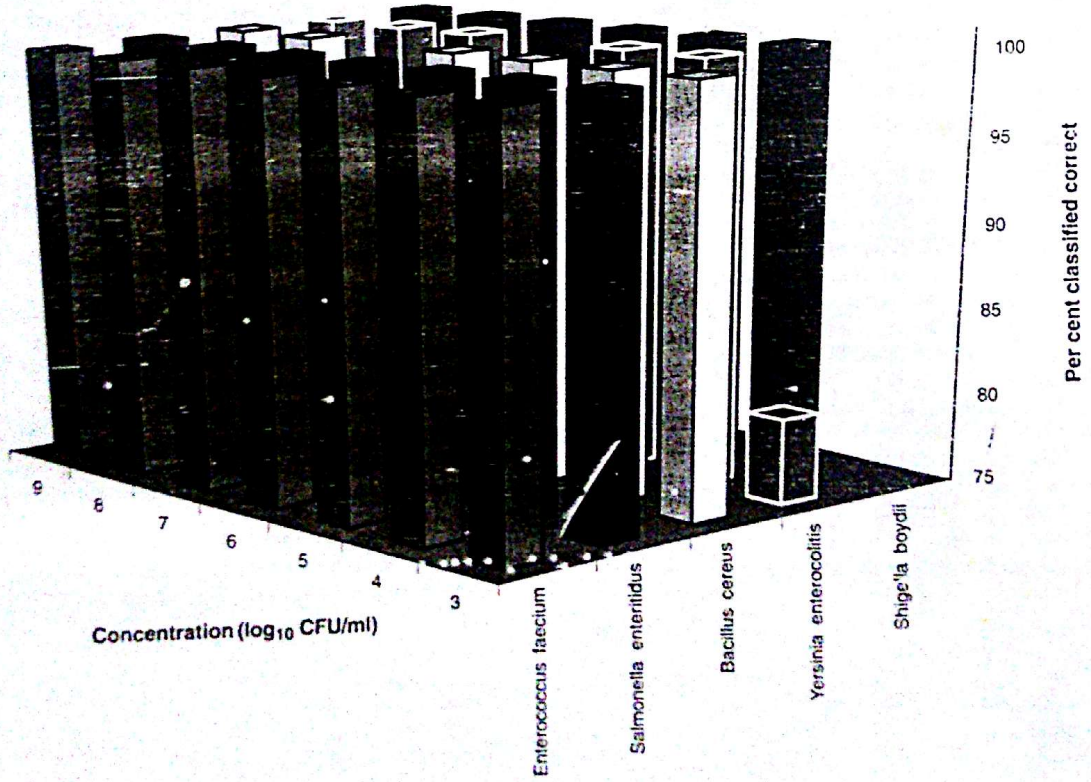

Figure 2. Classification of the bacterial species by ANN.

10^5, 10^6, and 10^9 CFU/mL concentrations; O30 at 10^6 CFU/mL; and O121 at 10^3 and 10^5 CFU/mL. No correlation was found between the number of wrongly classified strains and the concentrations. Thus, the error could only be attributed to the outlier patterns. This error could be reduced by increasing the size of the training set, which could be done only if the number of replications is increased.

Conclusion

ANNs were used for classification of bacteria at both the generic and serogroup levels based on their FTIR absorbance spectra. It was found that the classification was best with the training set containing 60% or more of the total data. The errors in the classification could be attributed to outlier patterns, which could be reduced by increasing the number of replications. Detailed analysis of the age of the culture and type of growth medium will provide additional insight into the sample variability. Interference due to water and carbon dioxide in the spectra could be reduced by proper purging and improved data collection, as found in newer FTIR spectrometers. The present approach could be extended to other food pathogens for generic, species, and serogroup level identification. Thus, FTIR spectroscopy with the use of ANNs could become a powerful tool for online food safety monitoring in the processing industry.

Acknowledgement

The authors gratefully acknowledge funding from the United States-Israel Binational Agricultural Research and Development Fund, Grant No. US-3296-02.

References

Argov, S., J. Ramesh, A. Salman, I. Sinelnikov, J. Goldstein, H. Guterman, and S. Mordechai. 2002. Diagnostic potential of Fourier-transform infrared microspectroscopy and advanced computational methods in colon cancer patients. *J. Biomedical Optics* 7(2): 1-7.

Chun, J., E. Atalan, A. C. Ward, and M. Goodfellow. 1993. Artificial neural network analysis of pyrolysis mass-spectrometric data in the identification of *Streptomyces* strains. *FEMS Microbiol. Letts.* 107(2-3): 321-325.

Crutchfield, S., P. Frenzen, J. Allshouse, and D. Roberts. 2000. Economics of food safety and international trade in food products. Presented at the *International Institute of Fisheries Economics and Trade IIFET 2000 Conference*. Corvallis, Ore.: Oregon State University.

Curk, M. C., F. Peladan, and J. C. Hubert. 1994. Fourier-transform infrared (FTIR) spectroscopy for identifying *Lactobacillus* species. *FEMS Microbiol. Letts.* 123(3): 241-248.

Freeman, R., R. Goodacre, P. R. Sisson, J. G. Magee, A. C. Ward, and N. F. Lightfoot. 1994. Rapid identification of species within the Mycobacterium tuberculosis complex by artificial neural network analysis of pyrolysis mass spectra. *J. Med. Microbiol.* 40(3): 170-173.

Goodacre, R., M. J. Neal, D. B. Kell, L. W. Greenham, W. C. Noble, and R. G. Harvey. 1994. Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs. *J. Microbiol. Methods* 76: 124-134.

Goodacre, R., E. M. Timmins, P. J. Rooney, J. J. Rowland, and D. B. Kell. 1996. Rapid identification of *Streptococcus* and *Enterococcus* species using diffuse reflectance-absorbance Fourier-transform infrared spectroscopy and artificial neural networks. *FEMS Microbiol. Letts.* 140(2-3): 233-239.

Goodacre, R., E. M. Timmins, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell, and P. J. Rooney. 1998. Rapid identification of urinary tract infection bacteria using hyperspectral, whole-organism fingerprinting and artificial neural networks. *Microbiology* 144(5): 1157-1170.

Helm, D., H. Labischinski, and D. Naumann. 1991a. Elaboration of a procedure for identification of bacteria using Fourier-transform IR spectral libraries: A stepwise correlation approach. *J. Microbiol. Methods* 14: 127-142.

Helm, D., H. Labischinski, H. Schallen, and D. Naumann. 1991b. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *J. Gen. Microbiol.* 137(1): 69-79.

Kirschner, C., N. A. Ngo Thi, and D. Naumann. 1999. Infrared imaging: An emerging tool for tissue diagnostics. In *Spectroscopy of Biological Molecules: New Directions*. 561. G. Greve, G. J. Puppels, and C. Otto, eds. Dordrecht. The Netherlands: Kluwer Academic Publishers.

Mead, P. S., V. Slutsker, L. F. Dietz, J. S. McCaig, S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe. 1999. Food-related illness and death in the United States. *Emerg. Infect. Dis.* 5(5): 607-625.

Naumann, D. 2000. Infrared spectroscopy in microbiology. In *Encyclopedia of Analytical Chemistry*, 102-131. R. A. Meyers, ed. Chichester, U.K.: John Wiley and Sons.

Naumann, D., D. Helm, and H. Labischinski. 1991a. Microbiological characterizations by FT-IR spectroscopy. *Nature* 351: 81-82.

Naumann, D., D. Helm, H. Labischinski, and P. Giesbrecht. 1991b. The characterization of microorganisms by Fourier-transform infrared spectroscopy (FT-IR). In *Modern Techniques for Rapid Microbiological Analysis*, 67-85. W. H. Nelson, ed. New York, N.Y.: VCH.

Naumann, D., S. Keller, D. Helm, C. Schultz, and B. Schrader. 1995. FT-IR spectroscopy and FT-Raman spectroscopy are powerful analytical tools for the non-invasive characterization of intact microbial cells. *J. Mol. Struct.* 347: 399-406.

Specht, D. F. 1990. Probabilistic neural networks. *Neural Networks* 3(1): 109-118.

Udelhoven, T., D. Naumann, and J. Schmitt. 2000. Development of a hierarchical classification system with artificial neural networks and FT-IR spectra for the identification of bacteria. *Appl. Spectrosc.* 54(10): 1471-1479.

Wallace, V. P., J. C. Bamber, D. C. Crawford, R. J. Ott, and P. S. Mortimer. 2000. Classification of reflectance spectra from pigmented skin lesions, a comparison of multivariate discriminant analysis and artificial neural networks. *Phys. Med. Biol.* 45(10): 2859-2871.

Yang, H., and J. Irudayaraj. 2003. Rapid detection of foodborne microorganisms on food surface using Fourier-transform Raman spectroscopy. *J. Mol. Struct.* 646: 35-43.