JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 1995, p. 2894–2898 Vol. 33, No. 11
0095-1137/95/$04.00+0
Copyright © 1995, American Society for Microbiology

Escherichia coli O157:H7 Restriction Pattern Recognition by Artificial Neural Network

C. ANDREW CARSON,[1]* JAMES M. KELLER,[2] KELLY K. McADOO,[1] DAYOU WANG,[2] BARBARA HIGGINS,[1] CRAIG W. BAILEY,[1] JAMES G. THORNE,[1] BEVERLEY J. PAYNE,[3] MAHREE SKALA,[3] AND ALLEN W. HAHN[4]

Departments of Veterinary Microbiology,[1] Electrical and Computer Engineering,[2] and Veterinary Clinical Sciences,[4] University of Missouri, Columbia, Missouri 65211, and Missouri Department of Health, Jefferson City, Missouri 65101-0570[3]

Received 8 May 1995/Returned for modification 14 July 1995/Accepted 9 August 1995

An artificial neural network model for the recognition of *Escherichia coli* O157:H7 restriction patterns was designed. In the training phase, images of two classes of *E. coli* isolates (O157:H7 and non-O157:H7) were digitized and transmitted to the neural network. The system was then tested for recognition of images not included in the training set. Promising results were achieved with the designed network configuration, providing a basis for further study. This application of a new generation of computational technology serves as an example of its usefulness in microbiology.

In recent years molecular typing methods have revolutionized the epidemiologic investigation of food-borne illnesses (17) by facilitating very specific identifications of suspected agents. The pulsed-field gel electrophoresis (PFGE) procedure (23), modified by contour-clamped homogeneous electric field technology (4), separates up to megabase-range fragments of genomic DNA after digestion with restriction enzymes. Restriction profiles of *Klebsiella pneumoniae* DNA, generated by PFGE, allowed the precise characterization of strains and isolates beyond the species level (7). A computer-assisted pattern recognition model has been used to identify mycobacteria to the species level on the basis of restriction fragment length polymorphism analysis (20).

Differences among phenotypically similar *Escherichia coli* O157:H7 (1) isolates have been detected by molecular means, and this form of fingerprinting has distinguished *E. coli* O157:H7 from nontoxigenic *E. coli* and Shiga-like-toxin-producing *E. coli* strains of other serogroups (3). Probe hybridization to Southern blots of various *E. coli* isolates indicated that various virulence-associated genes were located on DNA fragments of different lengths (18). This analytical means has been described as highly discriminatory for detecting genomic differences among strains (8).

The present study was designed to test the usefulness of an advanced form of computer-based pattern recognition which in this case was adapted to bacterial discrimination and identification. Basic pattern scanning is currently available via manual optical imaging, and pattern comparison can be done by expensive software. Our intention in this study was to design a computer-based decision algorithm to accomplish automated discrimination (19) at a modest cost. Conventional computer systems with sequential processors and distinct memory units are not equipped to handle a large amount of information that is fuzzy, probabilistic, noisy, inconsistent, or mutually interacting (5, 16). Artificial neural network (ANN) models, i.e., software composed of elements known as neurons or nodes and interconnections known as synapses or weights, have been proposed as alternatives (9, 10). Such systems have been used for the analysis of pyrolysis mass spectra for identification of *Mycobacterium tuberculosis* complex species (6) and prediction of *E. coli* promoter sites (11). ANN models store information in the strength of connections between nodes, which is adjusted to improve performance. This process is called adaptation, learning, or training. Among many ANN models, the multilayer feedforward network model is the most popular for a variety of applications and was chosen for this work. Our intent was to determine how such a model would perform when an ordinary personal computer and inexpensive software were used.

MATERIALS AND METHODS

Samples. One hundred one isolates of *E. coli* O157:H7 and 99 isolates of non-O157:H7 *E. coli* were analyzed. Samples, collected over a period of several years, were provided by the Missouri, Texas, Kansas, Utah, and Pennsylvania Departments of Health and by Health Canada. The Canadian samples were specifically selected to include isolates from various parts of the country. All isolates were previously identified by standard biochemical means and serology and were reconfirmed by similar methods in our laboratory. All O157:H7 samples were determined to be the same ribotype. Identified phage types included 1, 2, 8, 14, 23, and 32. Most isolates were derived from human diarrhea stool specimens, but some animal isolates were also included in the study (Table 1).

Preparation of bacterial DNA. All isolates were cultured to log phase in brain heart infusion broth. An aliquot (1.3 ml of cell suspension) of each isolate was centrifuged at 12,000 × *g* for 90 s, the supernatant was removed, and the pellet was washed twice in 1.0 ml of SE (75 mM NaCl, 25 mM EDTA, pH 8.0) buffer under the same centrifugation conditions. The optical density at 610 nm of the suspension was adjusted to 1.2 with SE diluent. The final suspension was mixed with an equal volume of warm 1% agarose prepared in modified TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0), poured into molds (10 by 15 mm), and allowed to solidify. Agarose plugs were incubated overnight at 50°C in lysis buffer (50 mM Tris [pH 8.0], 50 mM EDTA [pH 8.0], 1% Sarcosine, 1 mg of proteinase K per ml). The plugs were rinsed briefly with deionized sterile water, TE buffer containing 15.0 μl of phenylmethylsulfonyl fluoride (17 mg/ml in isopropanol) per ml was added to tubes containing the plugs, and the samples were incubated at room temperature for 30 min. After a second wash with TE buffer containing phenylmethylsulfonyl fluoride, the blocks were again rinsed briefly with sterile deionized H_2O, and the final four washes were in TE buffer for 30 min each. The agarose plugs were stored in TE buffer at 4°C.

Restriction digests. Plugs were cut into slices to fit the gel wells and equilibrated twice at 15-min intervals in tubes containing 200 μl of restriction enzyme buffer (Stratagene, La Jolla, Calif.). The buffer was removed, and 15 U of *Xba*I with specific buffer was added to the tubes containing the gel plugs. Samples were incubated for 4 h at 37°C, and the enzyme solution was replaced by 0.5× TBE

* Corresponding author. Mailing address: WHO Collaborating Center for Enteric Zoonoses, 104 Connaway Hall, College of Veterinary Medicine, University of Missouri, Columbia, MO 65211. Phone: (314) 882-6550. Fax: (314) 884-5050. Electronic mail address: vmandyc@vetmed.vetmed.missouri.edu.

TABLE 1. *E. coli* isolates used in this study

Class and source of isolates	No. of isolates
O157:H7	
Animal	28
Human	73
Total	101
Non-O157:H7	
Animal	15
O26	6
O111	1
O69	1
Other	7
Human	84
Total	99
Total	200

(0.089 M Tris-borate, 0.002 M EDTA) buffer and equilibrated for at least 10 min prior to electrophoresis.

PFGE. A 1.0% SeaKem (FMC, Rockland, Maine) gel was prepared in 0.5× TBE buffer approximately 1 h before completion of the restriction digestion of bacterial DNA and allowed to solidify at room temperature. Enzymatically digested DNA in gel slices was placed into the wells of the SeaKem gel and sealed in place with warm 1% agarose in TBE buffer. After the sealing agar solidified, the gel was transferred into the electrophoresis chamber (CHEF DRII or DRIII; Bio-Rad, Hercules, Calif.) and submerged in chilled 0.5× TBE. Electrophoresis was performed at 200 V for 20 h with the pulse time ramped from 5 to 50 s and the buffer temperature maintained at 14°C. The gel was stained with an ethidium bromide (0.5 μg/ml) solution for approximately 30 min at room temperature and then destained in distilled water for a minimum of 1 h prior to photography under UV light.

Digital imaging. Polaroid photographs of PFGE patterns were recorded by a desktop scanner (Hewlett-Packard, Boise, Idaho). Two separate processes were used to accomplish feature extraction. The first method involved a preprocessing function, performed on the entire pattern profile, to detect and digitize the first 10 bands (peaks) along the medial axis of the trace from the level of the largest marker band. The two parameters associated with each peak are normalized height and relative location, with respect to the trace origin. The location was normalized within the context of 0 and 1, representing the trace length compared with the marker lane. The pattern contains 261 pixels and is normalized between 0 and 255, with the lowest point equal to zero. Figure 1 represents the profile extracted from an O157:H7 lane pattern. A generic imaging software program (NIH Image), which allows calls from other software programs, was used to generate the profiles (Fig. 1A) along the medial axes of patterns. The profiles were used to automatically determine peaks (Fig. 1B) representing scanned bands located in the range between 100 kb and the largest marker band. The design of the second method of feature extraction was based on the presence of a major band at approximately 300 kb which was frequently a part of the O157:H7 pattern profile. By using a sliding window of one-sixth of the profile length, the band region with the highest summed density was determined. The regions above and below the most intense segment were then subdivided to generate six band regions. The extracted features became the midpoint location and average intensity of each of the six regions. The features generated by each extraction method were input to the ANN separately. Since many feature vectors for the two classes (O157:H7 and non-O157:H7) are very similar, crisp desired outputs were not assigned to the training vectors. Instead, fuzzy labels (2) were used to indicate the degree to which the features resemble other vectors from each class. Fuzzy memberships were assigned by using a modified version of the fuzzy K–nearest-neighbor algorithm (12).

Neural network software. The neural network architecture (Fig. 2) consisted of a multilayer feedforward network with 20 input nodes, 50 hidden nodes, and two output nodes (one for each class). The ANN was "trained" to relate each set of features, extracted by both methods described above, to a distinct class (O157:H7 or non-O157:H7) designation by using a back-propagation algorithm (21). Pattern recognition involves the estimation of a function that assigns a class label to the PFGE image. The multilayer feedforward network is used to assign a value of between 0 and 1 for each class of *E. coli*, indicating the degree to which the input pattern represents that class. The nodes in each layer are in contact through weighted interconnections that undergo modification during the "learning" process, resulting in pattern recognition.

Computer analysis and identification of PFGE patterns. PFGE patterns are scanned into computer memory as a digital image. To accomplish class distinction, i.e., *E. coli* O157:H7 or non-O157:H7, the ANN was trained to learn the distinctive features of *E. coli* O157:H7 profiles and determine to which class an unknown image belongs. Images presented during the learning phase were designated either O157:H7 or non-O157:H7.

Pulsed-field gels, containing a total of 101 *E. coli* O157:H7 isolates and 99 non-O157:H7 isolates, were analyzed. Ninety-one *E. coli* O157:H7 profiles and 89 non-*E. coli* O157:H7 profiles were used in each training iteration. For each subsequent interrogation, the remaining O157:H7 and non-O157:H7 profiles were input to the network as a "leave-20-out" strategy. This training-testing process was iterated 100 times so that in each instance 20 different isolates were used for testing; thus, each isolate was used "blindly" in 10 of the test iterations. The average of the ANN outputs of the two sets of input features was used to determine class recognition. Statistical analyses of sensitivity, specificity, positive predictive value, and negative predictive value were performed to evaluate the data generated (22).

A concomitant exercise based on human visual examination of all restriction patterns was performed to compare the accuracy of the computer with the recognition skills of two laboratory technicians who were experienced in the analysis of PFGE patterns. The technicians, who had seen all of the isolate profiles previously, were asked to classify the entire pattern profiles, which were presented blindly for their observation. The comparative performances of the ANN and the technicians were determined by calculation of pre- and posttest probability values for *E. coli* O157:H7 classification (22).

RESULTS

PFGE. Pulsed-field profiles of *E. coli* isolates were composed of distinct patterns with well-resolved bands (Fig. 1). Variations in the protocol led to differences in the patterns

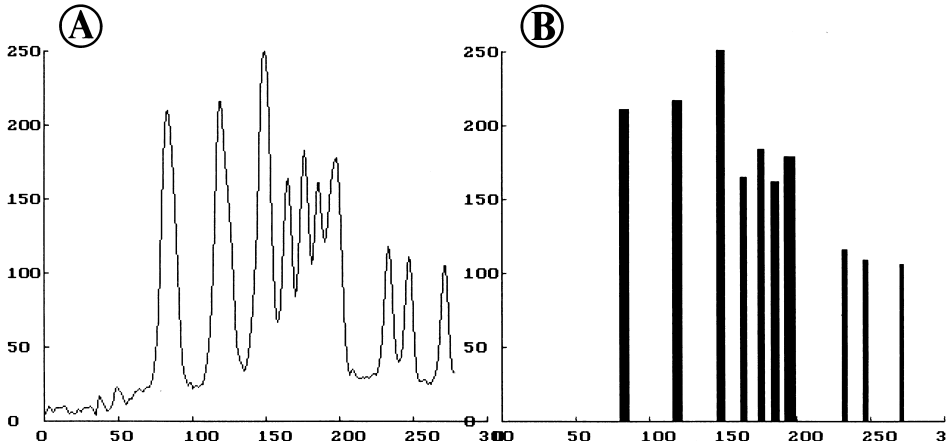

FIG. 1. Computer-generated image profile of PFGE pattern (A) and corresponding peaks extracted to minimize data (B). Values on the vertical axis represent relative band intensities; values on the horizontal axis represent relative band distances from the point of origin, the level of the largest marker band.

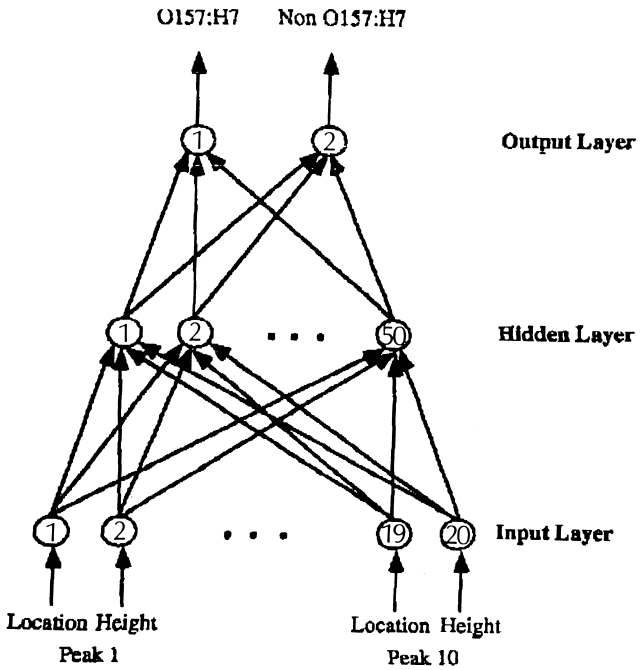

FIG. 2. Schematic of three-layer ANN model. Pattern features become inputs to the network at the input node layer. Numerical values for features are computed at noninput nodes; outputs of nodes are functions of weighted summations of values from units in the previous layer. Each sequential pair of the 20 input nodes records both the locations and heights of the first 10 recorded peaks (bands).

generated for individual isolates which affected subsequent analysis. However, the necessary reproducibility, in the form of the nearly identical patterns essential for ANN analysis, was achieved by strict adherence to prescribed gel protocols, quality control of large reagent batches, and very specific photographic conditions. Early recognition of operator-related differences indicated the additional requirement that multiple operators perform procedures in an identical manner.

Test of neural network ability to detect *E. coli* O157:H7. Numerous ANN designs, with and without data preprocessing functions, were tested and compared for accuracy. Data normalization and preprocessing were found to dramatically improve ANN performance. With the leave-10-out (for each class) approach to testing, the ANN correctly identified 98 of 101 *E. coli* O157:H7 images (Fig. 3A) and 90 of 99 non-O157:H7 images (Fig. 3B). These scores represented averages of the two ANN outputs, one for each set of extracted features. The software program scored 3% false negatives and 9% false positives. This level of performance essentially represents a 97% sensitivity in recognition of the O157:H7 images and a 90 or 91% specificity in recognition of the non-O157:H7 images (Table 2).

TABLE 2. Classification of *E. coli* by an ANN

ANN class	No. of isolates		
	O157:H7	Non-O157:H7	Total
Positive	98	9	107[a]
Negative	3	90	93[b]
Total	101[c]	99[d]	200

[a] Positive predictive value, 98/107 = 0.916.
[b] Negative predictive value, 90/93 = 0.968.
[c] Sensitivity, 98/101 = 0.970.
[d] Specificity, 90/99 = 0.909.

For human pattern analysis, performed for comparison with the ANN, we initially asked technicians to base their conclusions on the entire restriction pattern of each isolate. Comments from personnel indicated that they learned to center their attention on roughly the first 80% of the pattern array. Frequent variations in the smaller fragments (below 100 kb) seemed to be misleading and inconsequential in the process of isolate comparison. Technician 1 classified 79 of 101 O157:H7 isolates correctly and 91 of 99 non-O157:H7 isolates correctly. This level of recognition represented a sensitivity of 78.2%, a specificity of 91.9%, and an overall accuracy of 170 of 200 or 85%. Technician 2 correctly classified 65 of 101 O157:H7 isolates and 89 of 99 non-O157:H7 isolates, representing 64.4% sensitivity, 89.9% specificity, and 77% overall accuracy. Moderate agreement between the technicians was reflected by the derivation of a kappa statistic of 0.50. The agreement between technician 1 and the ANN was 0.62, and that between technician 2 and the ANN was 0.49. The kappa measurement (ranging between 0.0 and 1.0, with 1.0 perfect) reflects agreement between individuals or test methods in excess of that expressed by chance (22).

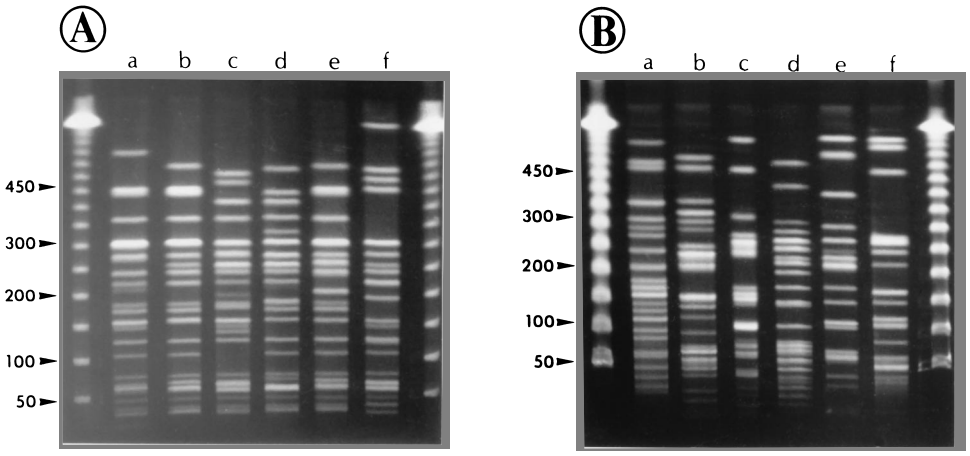

FIG. 3. PFGE patterns of *E. coli* O157:H7 isolates (A) and non-O157:H7 isolates (B). The outside lanes of each panel show size markers (in kilobases).

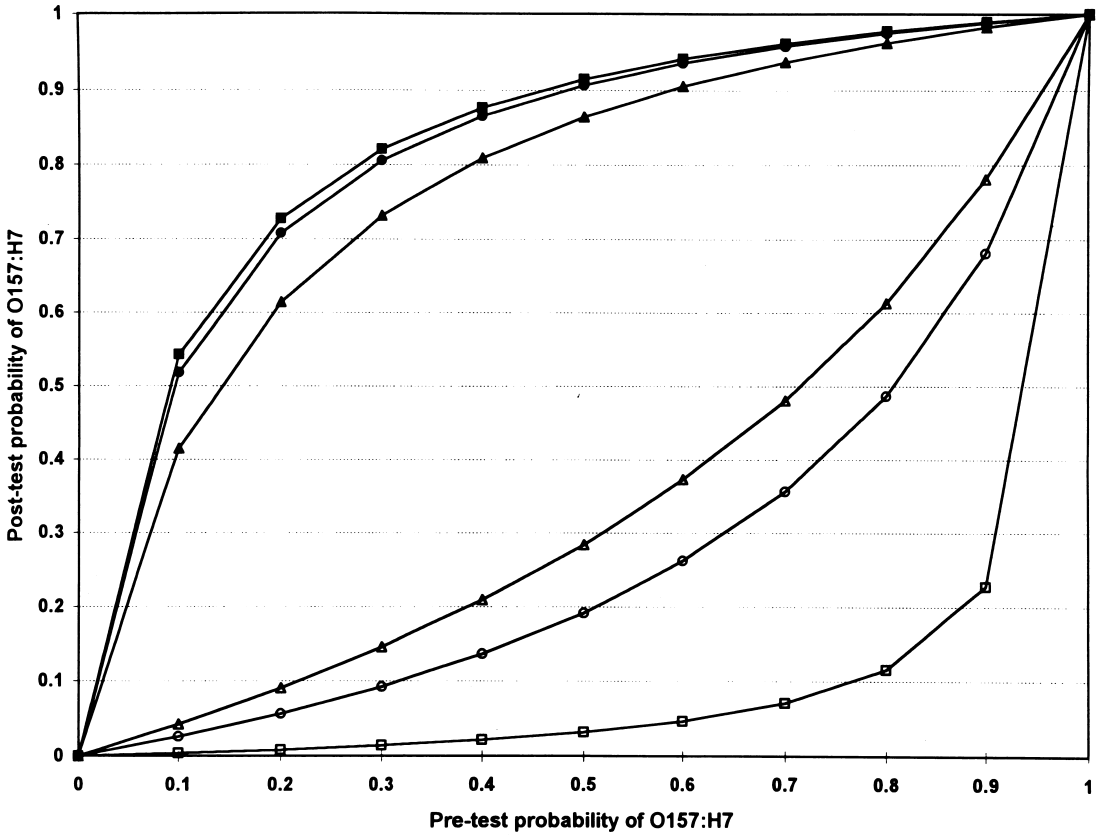

FIG. 4. Comparison of probabilities for classification of *E. coli* O157:H7 by the ANN and two technicians. Closed symbols, positive classification; open symbols, negative classification; squares, ANN; circles, technician 1; triangles, technician 2.

Comparisons of posterior probabilities for the technicians and the ANN are depicted in Fig. 4. The graphs correspond to the likelihood that a DNA profile is actually *E. coli* O157:H7, given a positive or negative test result (22). The uppermost lines (representing the ANN and technician 1 in classifying a pattern as *E. coli* O157:H7) are almost superimposed, indicating the best (and almost identical) performances. The ANN recognition of non-O157:H7, represented by the bottommost trace, was singularly best in this category. The ANN positive predictive value was 91.6%, and the negative predictive value was 96.2% (Table 2).

DISCUSSION

Numerous software programs successfully perform pattern analysis, clustering, and comparison, but the ANN is uniquely designed to render a "learned decision-making" function. We examined the task of identification of *E. coli* O157:H7 as an approach to the application of ANN technology, and the results indicate that recognition of microorganism patterns by ANN can be accomplished with a high level of accuracy. The described network training and testing exercises were accomplished with only a modest number of samples, and it is important to note that performance improved as sample size increased. The package described here is not commercially available. It is still in the developmental stage but could be made available for use, especially by reference laboratories, after additional modification and testing and publication of the final structure.

By comparison with human visual inspection, our ANN model appeared to process decision parameters in a superior fashion. It is not possible to determine the parameters upon which ANN decisions are based, but they could relate to the recognition of individual and relative fragment positions similar to those described by the technicians as typical for O157:H7 patterns, including heavy bands at 450 and 300 kb and four or five bands clustered between 300 and 200 kb (Fig. 3A). The ANN positive predictive value represents the proportion of test-positive samples that are actually O157:H7. The negative predictive value represents the proportion of ANN test-negative samples that are truly negative. Superior test systems are reflected in the probability graph (Fig. 4) by greater curvature above and below a diagonal from the origin. Predictive values are dependent on the prevalence (53.5% in this study), and with increasing prevalence the positive predictive value increases and the negative predictive value decreases. The opposite is true for decreasing prevalence. Therefore, if an isolate has a higher (>53.5%) probability of being O157:H7, the positive predictive value would be higher than that found in the present study. If the isolate has a lower (<53.5%) probability of being O157:H7, the positive predictive value would be less than that reported for the present ANN results.

There are numerous strategies which may be employed to further improve the computer-based analytical process. The described method of preparation for PFGE plus the run time requires approximately 4 days from sample receipt to final classification. We have recently shown that modifications can shorten the process to 3 days. Clustering algorithms can be used to combine training patterns into groups, thereby establishing subclass categories of both O157:H7 and non-O157:H7

samples. Multiple output nodes may then be designed to allow the computer program a wider range of class possibilities. Fuzzy logic can be applied to this phase of the system by implementation of fuzzy classes representing a "probable or undecided" category (2, 13). Isolates so designated could be subjected to further test analysis (e.g., genomic patterns produced by another method) or presented to a second-stage ANN. There are numerous connections between fuzzy logic and neural networks (2, 14, 15), including trainable network structures which implement fuzzy set theoretic operations. These structures explicitly model the uncertainty present in the data and could be used to improve the overall classification.

We recognize the natural extensions and ramifications of ANN-based pattern analysis. The software described in this report was not designed to address functions such as the identification of differences or similarities between isolate patterns, which is valuable information for epidemiology. Work (unpublished) in our laboratory indicates that the PFGE method, using *Xba*I, is a very specific means of fingerprinting for this approach to isolate identification. Several *E. coli* strains tested had nearly identical patterns, and a subsequent check of patient records confirmed similarities, substantiating the probability of the occurrence of an outbreak. In this regard, an ANN model could be designed specifically to detect outbreaks by class recognition. Very costly commercial software is currently available for the performance of such functions, but an ANN-based system of pattern analysis can be assembled in most laboratories for $10,000 to $20,000.

The described method of image recognition may be applicable to a wide range of microorganisms. The ANN model could be combined with a clustering algorithm so that various genera (classes) could translate to respective output nodes. The ANN model may also be applied to any preparative method that generates "fingerprint-like" patterns, such as random amplified polymorphic DNA analysis. Initial studies on the comparison of these methods of pattern generation in our laboratory indicated that PFGE profiles are superior to random amplified polymorphic DNA analysis in the establishment of clusters established on the basis of similarity. The ultrasensitive random amplified polymorphic DNA analysis procedure tends to magnify differences between isolates, making clustering difficult. We therefore submit that the pulsed-field method yields an appropriate level of pattern complexity for ANN recognition.

ACKNOWLEDGMENTS

This work was supported by the University of Missouri Research Board.

E. coli samples were kindly provided by the Missouri Department of Health, Jefferson City; the Veterinary Medical Diagnostic Laboratory, Columbia, Mo.; Health of Animals Laboratory, Agriculture Canada, Guelph, Ontario, Canada; the World Health Organization *E. coli* Reference Center, Pennsylvania State University, University Park; the Center for Control and Prevention of Communicable Diseases, Atlanta, Ga.; and the U.S. Department of Agriculture Food Safety Inspection Service, Beltsville, Md. We greatly appreciate the expert word-processing services rendered by Ellen Swanson of the Department of Veterinary Microbiology.

REFERENCES

1. **Arbeit, R. D., M. Arthus, R. Dunn, C. Kim, R. K. Selander, and R. Goldstein.** 1990. Resolution of recent evolutionary divergence among *Escherichia coli* O157:H7 from related lineages: the application of pulsed field electrophoresis to molecular epidemiology. J. Infect. Dis. **161:**230–235.
2. **Bezdek, J., and S. Pal (ed.).** 1992. Fuzzy models for pattern recognition. IEEE Press, New York.
3. **Bohm, H., and H. Karch.** 1992. DNA fingerprinting of *Escherichia coli* O157:H7 strains by pulsed-field gel electrophoresis. J. Clin. Microbiol. **30:** 2169–2172.
4. **Chu, G., D. Vollrath, and R. W. Davis.** 1986. Separation of large DNA molecules by contour-clamped homogeneous electric field. Science **234:** 1582–1585.
5. **Feldman, J. A., and D. H. Ballard.** 1992. Connectionist models and their properties. Cognitive Sci. **6:**205–54.
6. **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:**170–173.
7. **Gouby, A., C. Neuwirth, G. Bourg, N. Bouziges, M. J. Carles-Nurit, E. Despaux, and M. Ramus.** 1994. Epidemiological study by pulsed-field gel electrophoresis of an outbreak of extended-spectrum β-lactamase-producing *Klebsiella pneumoniae* in a geriatric hospital. J. Clin. Microbiol. **32:**301–305.
8. **Harsano, K. D., C. W. Kaspar, and J. B. Luchansky.** 1993. Comparison and genomic sizing of *Escherichia coli* O157:H7 isolates by pulsed-field gel electrophoresis. Appl. Environ. Microbiol. **59:**3141–3144.
9. **Hecht-Nielsen, R.** 1989. Neurocomputing. Addison-Wesley Publishing Co., Redwood City, Calif.
10. **Hertz, J., A. Grogh, and R. G. Palmer.** 1991. Introduction to the theory of neural computation. Addison-Wesley Publishing Co., Redwood City, Calif.
11. **Horton, P. B., and M. Kanehisa.** 1992. An assessment of neural network and statistical approaches for prediction of *E. coli* promoter sites. Nucleic Acids Res. **20:**4331–4338.
12. **Keller, J., M. Gray, and J. Givens.** 1985. A fuzzy K nearest neighbor algorithm. IEEE Trans. Syst. Man Cybernet. **15:**580–585.
13. **Keller, J., and R. Krishnapuram.** 1994. Fuzzy decision models, p. 213–232. *In* R. Yager and L. Zadeh (ed.), Computer vision in fuzzy sets, neural networks and soft computing. Van Nostrand, New York.
14. **Keller, J., Krishnapuram, R., Chen, Z. and Nasraoui, O.** 1994. Fuzzy additive hybrid operators for network-based decision making. Int. J. Intelligent Syst. **9:**1001–1024.
15. **Keller, J., R. Yager, and H. Tahani.** 1992. Neural network implementation of fuzzy logic. Fuzzy Sets Syst. **45:**1–12.
16. **Lippmann, R. P.** 1987. An introduction to computing with neural nets, vol. 4, p. 4–22. IEEE Press, Piscataway, N.J.
17. **Motalebi, M. S.** 1990. Molecular epidemiology of Shiga-like toxin producing *Escherichia coli* in food. Ph.D. dissertation. University of Washington, Seattle.
18. **Ott, M., L. Bender, G. Blum, M. Schmittroth, M. Achtman, H. Tschape, and J. Hacker.** 1991. Virulence patterns and long-range genetic mapping of extraintestinal *Escherichia coli* K1, K5, and K100 isolates: use of pulsed-field gel electrophoresis. Infect. Immun. **59:**2664–2672.
19. **Piper, J., and E. Granum.** 1989. On fully automatic feature measurement for banded chromosome classification. Cytometry **10:**242–255.
20. **Plikaytis, B. D., B. B. Plikaytis, and T. M. Shinnick.** 1992. Computer-assisted pattern recognition model for the identification of slowly growing mycobacteria including *Mycobacterium tuberculosis*. J. Gen. Microbiol. **138:**2265–2273.
21. **Rumelhart, D. E., and J. McClelland.** 1986. Parallel distributed processing. Explorations in the microstructure of cognition, vol. 1. Foundations. MIT Press, Cambridge, Mass.
22. **Sackett, D. L., R. B. Haynes, G. H. Guyatt, and P. Tugwell.** 1991. Clinical epidemiology, 2nd ed., ch. 1, p. 55–99. Little, Brown and Company, Boston.
23. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell **37:**67–75.