J. F. Casale,[1] *B.S. and J. W. Watterson,*[2] *Ph.D.*

A Computerized Neural Network Method for Pattern Recognition of Cocaine Signatures

REFERENCE: Casale, J. F. and Watterson, J. W., "**A Computerized Neural Network Method for Pattern Recognition of Cocaine Signatures,**" *Journal of Forensic Sciences*, JFSCA, Vol. 38, No. 2, March 1993, pp. 292–301.

ABSTRACT: This article describes a practical procedure for rapidly searching a large database of cocaine signatures to identify database entries that closely resemble a given reference cocaine exhibit using a personal computer (PC). The procedure takes advantage of the pattern recognition capability of the multilayer perceptron neural network to identify similar cocaine signatures. A PC-based software implementation is now being used on a daily basis at the North Carolina State Bureau of Investigation (NCSBI) to aid forensic experts in identifying signatures that originate from the same batch. Intelligence reports generated from database searches have been useful to undercover agents in the field who are striving to build drug related conspiracy cases. This software was developed as a collaborative effort between the NCSBI and the Center for Systems and Engineering of the Research Triangle Institute.

KEYWORDS: criminalistics, cocaine, cocaine signatures, neural network models, illicit drugs

Chromatographic signature analyses of forensic evidence is typically carried out by visually comparing the corresponding components of signatures exemplified in Fig. 1. This approach to manual pattern matching is both time consuming and inaccurate for comparison of an exhibit with all entries in a large database. Along with the recent interests in chromatographic signature analyses of cocaine [*1–9*], interest in a reliable computerized pattern recognition program has emerged in several disciplines [*10–13*]. An excellent review of neural networks in forensic science is presented by Kingston [*13*].

When searching for a suitable pattern-recognition technique for comparing cocaine signatures it is important to recognize that cocaine signature matching is a very difficult statistical problem. Each component of a cocaine signature represents an impurity (percent by weight) found in the cocaine sample, and the corresponding components derived by repeated sampling from the same batch can vary slightly from one sample to the next. The problem is further complicated by the fact that some of the impurities are correlated because certain impurities are derived from other impurities [*10*]. Unfortunately, the statistical variation (or noise) that is associated with each impurity is not known at the present time. In this regard, an optimum cocaine signature matching technique is one

Received for publication 4 April 1992; revised manuscript received 10 July 1992; accepted for publication 13 July 1992.

This work was supported in part by grants 170-188-E6-D009 and 170-190-E6-D034 from the U.S. Department of Justice.

Presented in part at the International Symposium on the Forensic Aspects of Trace Evidence, 24–28 June 1991, Quantico, VA.

[1]Forensic Chemist, DEA Special Testing, McLean, VA.

[2]Senior Research Engineer, Research Triangle Institute, Research Triangle Park, NC.

that maximizes probability of identifying signatures from the same batch while minimizing the probability of false positives.

The ultimate goal of signature profiling cocaine exhibits is to build a computerized database of samples to be rapidly searched against new exhibits for possible matches. The search results are to be used only as an aid to the chemist. We have found that a database coupled with a computerized search capability has enhanced exhibit correlations in narrowing a broad field of possible matches for the chemist. Thus, a timely output of intelligence information is disseminated to the law enforcement community and for use in building conspiracy cases.

Known approaches for achieving cocaine signature classification include distance functions (nearest neighbor, clustering, etc.), statistical (correlation, maximum likelihood, etc.), and neurocomputing (visual observation, neural networks, etc.). All three approaches were evaluated as a comparison tool. Commercial software programs were used to evaluate distance functions and statistical approaches, and it was observed that they were inherently slow, or produced unacceptable classification results, or both. Alternatively, our research revealed that the layered neural network is capable of rapidly comparing cocaine signatures and indicating which database signature most closely resembles a new cocaine exhibit.

Continuing the research effort, we have developed a neural network software tool that aids the forensics expert in identifying chromatograms of cocaine samples from the same batch. The following sections describe the neural network and the software implementation now being used on a daily basis to automatically search a database and identify

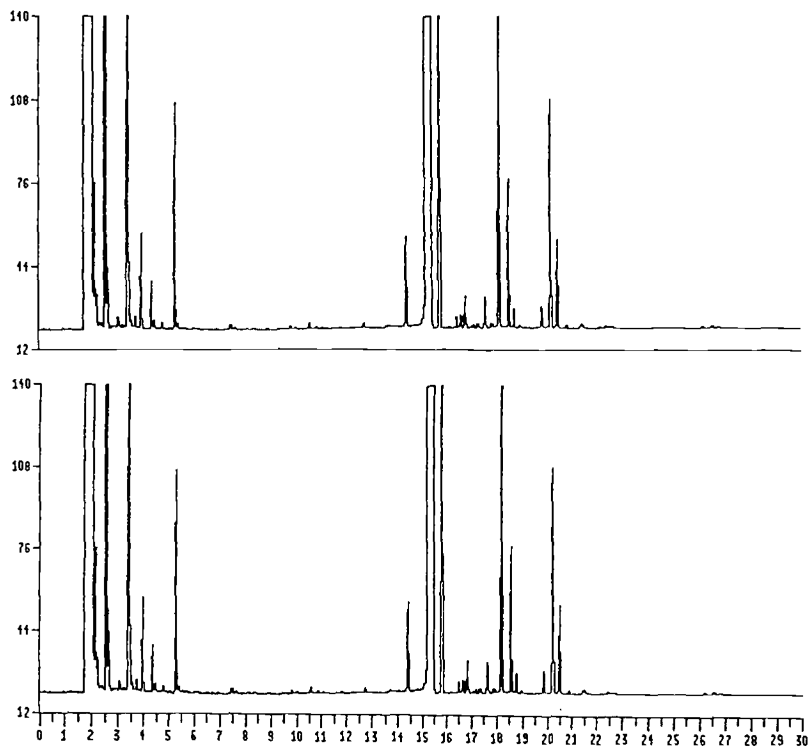

FIG. 1—*Similar cocaine signatures.*

illicit cocaine samples containing trace amounts of coca related impurities that closely resemble a given reference exhibit.

Neural Network Description

Neural networks were proposed more than forty years ago as a model of the human brain [*14*]. Over the years these networks have been studied in an attempt to achieve humanlike performance in the fields of speech and image recognition. The attractiveness of neural networks is related to their special abilities in the areas of pattern classification, robustness, and real-time performance that have been demonstrated in simple neural network systems [*15*]. Neural networks are especially good at solving complex pattern-recognition problems implicit in understanding continuous speech, identifying handwritten characters, and determining that a target seen from different angles is in fact one and the same object. It is not surprising that neural networks are also useful for comparing cocaine impurity signatures and helping the forensics expert to identify cocaine signatures from the same batch.

Our laboratories demonstrated in an early phase of this research that the multilayer perceptron neural network [*16,17*] shown in Fig. 2, can be used to compare cocaine impurity signatures in a time efficient manner. The multilayer perceptron, the most widely used neural network model for solving pattern-recognition problems, was thus chosen for the cocaine signature classification problem. This network is a massive parallel processor that can process data at a high rate of speed when compared to alternative processing techniques. Regarding the detection capability of the neural network, an optimum multilayer perceptron neural receiver for detecting the presence of a signal corrupted by a band limited Gaussian noise is described in [*18*].

The network shown in Fig. 2 consists of nodes and interconnecting arcs that form signal paths from left to right through the network. Associated with the arcs are weights that are adjusted during training to allow the network to recognize training set patterns, and ultimately classify a given unknown exhibit. Each node in the chosen architecture is further defined in Fig. 3. Here it is observed that the sum of the weighted signal input $\{W_0X_0, W_1X_1, \ldots, W_{N-1}X_{N-1}\}$ and a bias parameter Θ_n are processed through sigmoid nonlinearity F_n to obtain the nth node output y_n given by

$$y_n = F_n(x) = \frac{1}{1 + e^{-x}}, \tag{1}$$

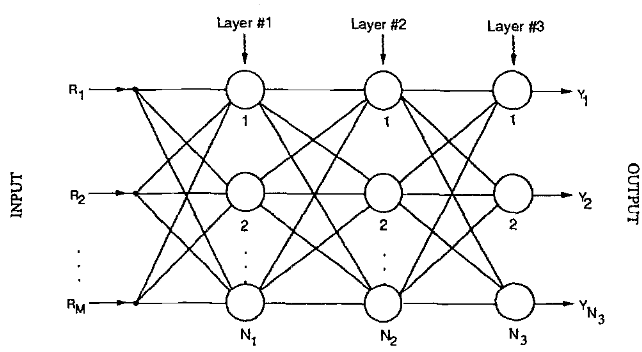

FIG. 2—*Multilayer perceptron neural network.*

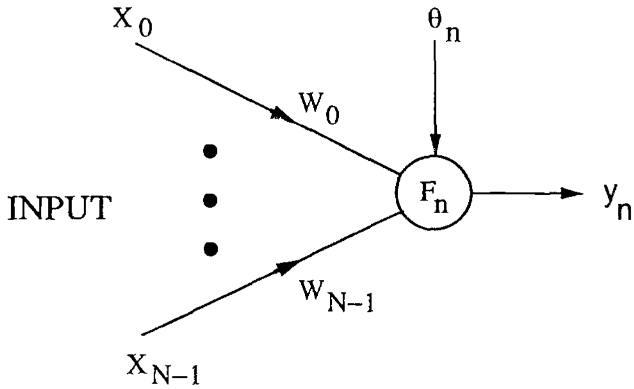

FIG. 3—*Chosen neuron model.*

where

$$x = \sum_{i=0}^{N-1} W_i X_i - \theta_n. \qquad (2)$$

Multilayer perceptron training is accomplished by repeated application of training patterns and the use of the back-propagation training algorithm to adjust the weights and node biases. This training technique is an interactive gradient algorithm designed to minimize the mean square error between the actual output of a multilayer feed-forward perceptron and the desired output. The training procedure requires the use of continuous differentiable nonlinearities such as the sigmoid function described in Eq. 1. A five-step back-propagation training algorithm is outlined in [*16*].

Experimental

The selection of the neural network training set was found to be critical to the success of the research effort. During early stages of this program, fruitless attempts were made to use the set of database patterns to train the network. Two problems that arose when using the set of database patterns to train the network were:

1. The neural network was very difficult to train when two or more database patterns were nearly the same.
2. When the network did train satisfactorily with the set of database patterns, the resulting pattern matching capability of the network was not acceptable. In retrospect this is not surprising since each database pattern typically comes from a different batch of cocaine, and one pattern is not representative of the associated batch of cocaine.

An alternative approach to training set selection that produced excellent results was to randomly generate training patterns in the vicinity of the reference exhibit, each randomly generated pattern being viewed as an exemplar pattern for a pseudo class of patterns. A detailed description to this approach to training set selection is described in [*19*].

Having demonstrated the attractiveness of the neural network for matching cocaine signatures, efforts were directed toward selecting a specific multilayer perceptron structure. Using cocaine signature data, numerous simulation experiments were carried out to establish the required number of nodes and arcs in the multilayer perceptron. Simulation results revealed that a fully connected, two-layer perceptron neural network with 16 nodes per layer is capable of near-optimum identification of database patterns that closely match a given reference exhibit. Noting that there are 16 output lines in the chosen neural network structure, the question must also be addressed as how to interpret the trained neural network output when applying database entries to the network. Various approaches were examined, and the approach that gave the best results was the use of mean-square-error as a basis for deciding which database entry is closest to a reference exhibit.

The resulting neural network simulation software is an interactive C-language program developed for use on a personal computer (PC). The compiled program occupies 140K bytes of hard disk. Noting that the neural network training is computationally intensive, the PC should contain at least a 80386 microprocessor (or equivalent), a math co-processor, 2 megabytes RAM, and MS-DOS ver. 3.3 or higher. The time required to train the chosen neural network typically ranges from 1 to 3 min on a 386/25Mhz PC, and an additional time of approximately 30 seconds is required to search a database of 2000 patterns and identify database patterns that are closest to the reference pattern.

Acquisition of Cocaine Signatures

Cocaine signatures (containing 16 quantitated cocaine related impurities) were acquired using the previously reported procedure of Casale and Waggoner [*1*]. Two additional recently reported coca impurities [*20,21*], N-benzoylnorecgonine methyl ester and N-acetylnorcocaine, have been incorporated into this assay. From this assay a master database (DBASE1) of over 3000 cocaine signatures has been accumulated along with smaller subsets comprised of four separate localized major metropolitan areas (Wake, Durham, Guilford, and Cumberland) in North Carolina. Quantitative data from typical cocaine signatures are represented in Table 1.

Pattern Recognition Experiments

The qualitative and quantitative data obtained from 3426 different cocaine signatures were compiled into a large database for pattern recognition to newly acquired reference signatures. Newly acquired reference signatures were also appended to the database prior pattern recognition experiments. All experiments were performed on an Everex Step Model 386\25 MHz personal computer (MS-DOS ver. 3.3) with 2 megabytes RAM and 80 megabytes hard disk. A 387 math coprocessor was incorporated to reduce the time required to train the neural network. The output of a typical computer run is shown in Table 2.

The following is a sample of four experiments from the many hundreds of experiments that were performed during development of the pattern matching software.

Experiment #1

Experiment Description—An exhibit, separated into five identical samples (replicate analysis data), was processed by the pattern recognition program to determine if matches would be indicated within the database.

Expected Results—No matches were anticipated in this experiment because none of the database signatures came from the batch of cocaine used to generate the five samples.

Actual Results—The five samples were processed by the computer program, one at a time, to determine if any of the database patterns closely matched the exhibit. The

TABLE 1—*Typical cocaine signatures.*

CASE	CHEMICAL IMPURITIES (BY WEIGHT)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
91-21649	.10	.13	.00	.02	.00	.21	.04	.81	.00	.06	.09	1.73	.05	1.15	.07	.00
91-21650	.12	.06	.00	.03	.00	.25	.00	.20	.00	.05	.07	1.97	.07	1.26	.00	.00
91-21694-1	.06	.05	.00	.03	.25	.17	.02	.72	.18	.07	.23	2.21	.00	1.73	.00	.00
91-21694-2	.07	.05	.00	.03	.30	.21	.02	.75	.18	.07	.13	2.23	.00	1.24	.00	.00
92-1514	.02	.08	.00	.00	.16	.07	.02	.24	.00	.00	.10	3.01	.00	1.58	.25	.00
92-2140	.32	.35	.00	.03	.02	.48	.00	.12	.00	.00	.04	1.99	.06	1.31	.00	.09
92-2770	.08	.09	.00	.00	.02	.08	.00	.18	.06	.00	.04	0.10	.15	0.03	.00	.00
92-3877	.04	.09	.00	.00	.05	.04	.05	.09	.00	.03	.02	4.23	.00	2.02	.00	.00
92-5797-2	.10	.06	.00	.02	.03	.15	.00	.08	.00	.00	.05	4.46	.00	2.98	.04	.00

computer program correctly indicated that none of the samples closely matched any of the database signatures.

Experiment #2

Experiment Description—Two signatures from the same batch (exhibits A and B) were compared by inserting exhibit B in the database and using exhibit A as a reference pattern when training the neural network.

Expected Results—The computer program should indicate a close match between exhibits A and B, while at the same time indicating that exhibit A does not match any other database pattern.

Actual Results—The computer program indicated that exhibits A and B were from the same batch. No other database pattern was observed to match exhibit A.

Experiment #3

Experiment Description—Five exhibits from the same batch, each exhibit consisting of five replicate samples, were trained on the neural network and tested against the database. All 25 samples were in the database.

TABLE 2—*SNIFFER output file.*

```
**********************************************************
*                                                        *
*   SNIFFER.1; SOFTWARE FOR MATCHING COCAINE SIGNATURES  *
*              (VERSION 1.0; 8/9/91)                     *
*            RTI 1991 (ALL RIGHTS RESERVED)              *
*                                                        *
**********************************************************
DATE:    Tue Mar 24 11:08:02 1992
NUMBER OF TRAINING ITERATIONS = 81
+++++++++++++++++++ DATABASE SELECTION ++++++++++++++++++
 (1) -> dbase1
 (2) -> wake
 (3) -> durham
 (4) -> guilford
 (5) -> cumber
 (6) -> ANOTHER DATABASE NAME??
 (7) -> EXIT
 (Enter 1, 2, 3, 4, 5, 6, or 7):  1
Enter desired number of OUTPUT PATTERNS (integer):  5
NUMBER OF DATABASE PATTERNS =  3426
++++++++++++++++++++++++++++++++++++++++++++++++++++++++
reference         =           0.0000       R92-4928-HCL
++++++++++++++++++++++++++++++++++++++++++++++++++++++++
ix, %error[ix]    =  1        0.0000       R92-4928-HCL (ALERT)
++++++++++++++++++++++++++++++++++++++++++++++++++++++++
ix, %error[ix]    =  2        0.5922           r91-3192
++++++++++++++++++++++++++++++++++++++++++++++++++++++++
ix, %error[ix]    =  3        0.9069            r90-930
++++++++++++++++++++++++++++++++++++++++++++++++++++++++
ix, %error[ix]    =  4        0.9530           r89-12011
++++++++++++++++++++++++++++++++++++++++++++++++++++++++
ix, %error[ix]    =  5        1.0601      R92-4053-1HCLC
++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

Expected Results—Each of the 25 samples should have one exact match and closely match the 24 replicate database signatures.

Actual Results—The program indicated 25 matches (including one exact match) for each of the 25 samples. No matches were indicated for another of the database patterns.

Experiment #4

Experiment Description—The database of 3426 cocaine signatures was searched against itself for any *exact* matches between differing exhibits.

Expected Results—Each of the signatures in the database was used as a reference pattern when searching the database. Every reference pattern should then find itself in the database.

Actual Results—The computer program correctly indicated that, for each of the 3426 signatures, there was one (and only one) exact match in the database.

Results and Discussion

The computerized pattern recognition program is used to compare cocaine signature patterns using the process indicated in Fig. 4. Cocaine from a seizure is sampled and analyzed using a gas chromatograph to obtain a new (reference) signature. Through the use of a trained neural network, the database is automatically searched to identify database signatures from previous drug seizures that closely match the reference signature. The forensic chemist then visually compares the chromatogram of the reference signature with the database signature(s) identified by the neural network to ascertain whether the signatures come from the same batch. In no instance should the chemist rely solely on the computerized search. The search is only to be used to narrow a broad field of possible matches. The discussion as to whether two or more exhibits are related rests on the experience and expertise of the chemist interpreting the signature chromatograms.

The neural network software was found to be far superior to all other known computational search programs. This software will search a new cocaine signature against a database of over 10 000 cocaine signatures in less than three minutes. We have refined the neural network pattern recognition software to an "on-line" operational intelligence program in the NCSBI Central Drug Chemistry Laboratory. This includes an auto-search capability of searching 100 or more new signatures against a database in a single session.

Cocaine signature analyses and the presented software are used at the NCSBI laboratory to achieve the following objectives:

1. To determine if two or more specifically requested cocaine samples are from the same batch. In these cases the software is used as an additional technique to supplement the chemist's conclusions on the signatures. A laboratory report is written for specific requests comparisons which states if the samples are from the same batch, different batches, or inconclusive analysis. In a six month period 50 laboratory reports were provided for cocaine conspiracy investigations (23 same batch, 20 different batches, and seven inconclusive).

2. To provide intelligence information to officers on cocaine samples that are routinely submitted to the laboratory. Currently the laboratory is performing cocaine signature work on all cocaine samples submitted from four major metropolitan areas because of the quantity of submissions from those areas. Intelligence signatures are searched against

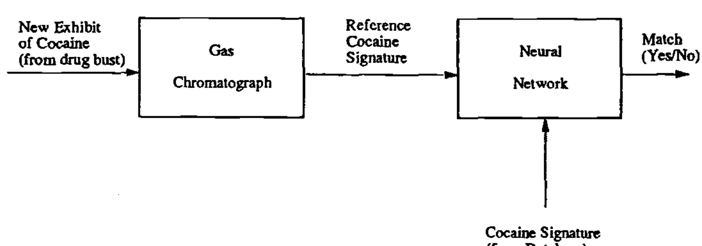

FIG. 4—*Cocaine signature matching procedure.*

each other in a large database. A significant number of the "hits" involve two or more law enforcement agencies. The officers are alerted through a Cocaine Signature Intelligence Memo stating the officers involved, agency, agency file number and Laboratory Report number. These intelligence reports have had a very positive impact on interagency corroboration and the development of conspiracy cases. Approximately 100 intelligence memos were provided in the sample six month period. It should be noted that these reports are intended for intelligence purposes only. We are currently running 300 to 500 intelligence signatures a month, and now have the capability to send quantitative cocaine signature data directly from the gas chromatograph data system to the neural network program, thus decreasing input/output time and eliminating data entry errors.

Conclusions

This article describes a collaborative effort that resulted in the development of a neural network pattern recognition software tool that can be used to rapidly search a large database and identify database entries that closely resemble a given cocaine exhibit's signature. The neural network was found to be far superior to distance functions classifications (nearest neighbor, clustering, etc.) and conventional statistical approaches (correlation, maximum likelihood, etc.). All chromatographic signature patterns from each experiment were correctly recognized and matched with no false positives or false negatives. Neural network pattern recognition was demonstrated to be a reliable means of rapidly establishing common source identity of cocaine samples. Results of this study using differing batches of cocaine may show potential as a pattern recognition method for processing "country of origin" determinations in combination with more sensitive chromatographic techniques using electron capture detection.

Acknowledgments

The authors wish to thank R. W. Waggoner and W. L. Cook for their help in acquiring many of the cocaine signatures used in this study.

References

[*1*] Casale, J. F. and Waggoner, R. W., "A Chromatographic Impurity Signature Profile Analysis for Cocaine Using Capillary Gas Chromatography," *Journal of Forensic Sciences*, Vol. 36, No. 5, September 1991, pp. 1312–1330.

[*2*] Lurie, I. S., Moore, J. M., and Kram, T. C., "Analysis of Manufacturing By-Products and Impurities in Illicit Cocaine via High-Performance Liquid Chromatography and Photodiode Array Detection," *Journal of Chromatography*, Vol. 405, 1987, pp. 273–281.

[*3*] Moore, J. M., Cooper, D. A., Lurie, I. S., Kram, T. C., Carr, S., Harper, C., and Yeh, J., "Capillary Gas Chromatographic-Electron Capture Detection of Coca-Leaf-Related Impurities in Illicit Cocaine: 2,4-Diphenylcyclobutane-1,3-Dicarboxylic Acids. 1,4-Diphenylcyclobutane-2,3-Dicarboxylic Acids and Their Alkalodial Precursors. The Truxillines," *Journal of Chromatography*, Vol. 410, 1987, pp. 297–318.

[*4*] Casale, J. F., "Detection of Pseudoecgonine and Differentiation from Ecgonine in Illicit Cocaine," *Forensic Science International*, Vol. 47, No. 3, October 1990, pp. 277–287.

[*5*] Lurie, I. S., Moore, J. M., Kram, T. C., and Cooper, D. A., "Isolation, Identification and Separation of Isomeric Truxillines in Illicit Cocaine," *Journal of Chromatography*, Vol. 504, 1990, pp. 391–401.

[*6*] LeBelle, M., Callahan, S., Latham, G., Lauriault, G., and Savard, C., "Comparison of Illicit Cocaine by Determination of Minor Components," *Journal of Forensic Sciences*, Vol. 35, No. 4, July 1991, pp. 1102–1120.

[*7*] Brewer, L. and Allen, A. C., "N-Formyl Cocaine: A Study of Cocaine Parameters," *Journal of Forensic Sciences*, Vol. 36, No. 3, May 1991, pp. 697–707.

[8] Ensing, J. G., Racamy, C., and deZeeuw, R. A., "A Rapid Gas Chromatographic Method for the Fingerprinting of Illicit Cocaine Samples," *Journal of Forensic Sciences*, Vol. 37, No. 2, March 1992, pp. 446–459.
[9] Jansen, K. E., Walter, L., and Fernando, A. R., "Comparison of Illicit Cocaine Samples," *Journal of Forensic Sciences*, Vol. 37, No. 2, March 1992, pp. 436–445.
[10] Casale, J. F. and Watterson, J. W., "A Neural Network Method for Pattern Recognition of Chromatographic Signature Patterns of Forensic Trace Evidence," *Proceedings of the International Symposium on the Forensic Aspects of Trace Evidence*, 24–28 June 1991, Quantico, Virginia, in press.
[11] Watterson, J. W., "Cocaine Impurity Signature Profile Analysis," *Research Report RTI/5019/01F*, Research Triangle Institute, December 1991.
[12] Long, J. R., Mayfield, H. T., and Henley, M. V., "Pattern Recognition of Jet Fuel Chromatographic Data by Artificial Neural Networks with Back-Propagation of Error," *Analytical Chemistry*, Vol. 63, No. 13, July 1991, pp. 1256–1261.
[13] Kingston, C., "Neural Networks in Forensic Science," *Journal of Forensic Sciences*, Vol. 37, No. 1, January 1992, pp. 252–264.
[14] Eberhart, R. C. and Dobbins, R. W., *Neural Network PC Tools, A Practical Guide*, Academic Press, Inc., San Diego, 1990.
[15] Roth, M. W., "Neural-Network Technology and Its Applications," *Johns Hopkins APL Technical Digest*, Vol. 9, No. 3, 1988, pp. 242–253.
[16] Lippmann, R. P., "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, April 1987, pp. 4–22.
[17] Rumelhart, D. E. and McClelland, J. L., *Parallel Distributed Processing*, MIT Press, Vol. 1, 1987.
[18] Watterson, J. W., "An Optimum Multilayer Perceptron Neural Receiver for Signal Detection," *IEEE Transactions on Neural Networks*, Vol. 1, No. 4, December 1990, pp. 298–300.
[19] Watterson, J. W., "A New Procedure for Selecting a Neural Network Training Set," Patent Disclosure, Research Triangle Institute, April 1992.
[20] Casale, J. F., "N-Acetylnorcocaine: A New Cocaine Impurity from Clandestine Processing I," *Journal of the Clandestine Laboratory Investigating Chemists Association*, Vol. 1, No. 4, October 1991, pp. 23–26.
[21] Ensing, J. G. and Hummelen, J. C., "Isolation, Identification, and Origin of Three Previously Unknown Congengers in Illicit Cocaine," *Journal of Forensic Sciences*, Vol. 36, No. 6, November 1991, pp. 1666–1687.

Address requests for reprints or additional information to
John F. Casale
DEA Special Testing and Research Laboratory
7704 Old Springhouse Rd.
McLean, VA 22102