

Detection and Classification of Organophosphate Nerve Agent Simulants Using Support Vector Machines with Multiarray Sensors

Omowunmi Sadik,^{*,†} Walker H. Land, Jr.,[‡] Adam K. Wanekaya,[†] Michiko Uematsu,[†]
Mark J. Embrechts,[§] Lut Wong,[‡] Dale Leibensperger,[‡] and Alex Volykin[‡]

Departments of Chemistry and Computer Science, State University of New York at Binghamton,
Binghamton, New York 13902-6000, and Department of Decision Sciences and Engineering Systems,
Rensselaer Polytechnic Institute, Troy, New York 12180-3590

Received October 8, 2003

The need for rapid and accurate detection systems is expanding and the utilization of cross-reactive sensor arrays to detect chemical warfare agents in conjunction with novel computational techniques may prove to be a potential solution to this challenge. We have investigated the detection, prediction, and classification of various organophosphate (OP) nerve agent simulants using sensor arrays with a novel learning scheme known as support vector machines (SVMs). The OPs tested include parathion, malathion, dichlorvos, trichlorfon, paraoxon, and diazinon. A new data reduction software program was written in MATLAB V. 6.1 to extract steady-state and kinetic data from the sensor arrays. The program also creates training sets by mixing and randomly sorting any combination of data categories into both positive and negative cases. The resulting signals were fed into SVM software for “pairwise” and “one” vs all classification. Experimental results for this new paradigm show a significant increase in classification accuracy when compared to artificial neural networks (ANNs). Three kernels, the S2000, the polynomial, and the Gaussian radial basis function (RBF), were tested and compared to the ANN. The following measures of performance were considered in the pairwise classification: receiver operating curve (ROC) A_z indices, specificities, and positive predictive values (PPVs). The ROC A_z values, specificities, and PPVs increases ranged from 5% to 25%, 108% to 204%, and 13% to 54%, respectively, in all OP pairs studied when compared to the ANN baseline. Dichlorvos, trichlorfon, and paraoxon were perfectly predicted. Positive prediction for malathion was 95%.

INTRODUCTION

The United States is coming under increasing threats of chemical warfare agents (CWAs) of mass destruction by international terrorist organizations and the new Department of Homeland Security is seeking new frontiers of technologies to combat these threats. Consequently, there is a great deal of interest in developing tools that can be used not only to detect but also to effectively classify CWAs. At the molecular level, all chemical warfare agents are strong electrophiles, containing either central phosphorus (e.g., Soman, Sarin, and Tabun), sulfur, or nitrogen (VX or mustards) atoms (Figure 1). The central atoms of the nerve agents (i.e., Sarin and Soman) attach to the ends of acetylcholinesterase enzyme and stay bound to it for many hours. Organophosphates (OPs) are potential CWAs because their action is very similar to nerve agents. OPs act by inhibiting the acetylcholinesterase enzyme, which is essential for functioning of the nervous system in humans. The inhibition of acetylcholinesterase results in the accumulation of acetylcholine that interferes with muscular responses that may be fatal.^{1,2} Early detection of OPs may give an indication of terrorist activity that may allow proper procedures to be followed to mitigate dangers.

Gas, liquid, and thin-layer chromatography coupled with different detectors are the most commonly used methods for the detection of OPs.^{3–5} However, these techniques, which are time-consuming and expensive and require highly trained personnel, are available only in sophisticated laboratories and are not amenable to on-line and rapid monitoring. Biological methods such as immunoassays, biosensors, and inhibition of cholinesterase activity for OP determination have also been reported.^{6–8} Immunoassays require long analysis time and extensive sample handling with multiple washing steps. Monitoring chemical and biological warfare agents (CBAs), residue in soil, water, food, and air is possible by screening or through diagnostic techniques that can provide only a qualitative “yes-or-no” answer, or semiquantitative/quantitative techniques, which can detect and quantify residues in the sub-threshold levels.⁹ It is possible for these methods to generate false positives or false negatives if the sensitivities are insufficient for the threshold levels. Applying any detection principle to a potential agent depends on the characteristics of the detection technique, the nature of the analytes, and the goal of the analysis system. Detectors that are designed for gas or vapor plumes may not readily be applicable to the detection of low volatile liquid, semivolatiles compounds, bacteria, or viruses. Thus, the need for fast responding and accurate CWA detection systems is expanding and the utilization of cross-reactive sensor arrays to detect volatile compounds in conjunction with pattern recognition techniques to interpret arrays response patterns may prove

* Corresponding author fax: (607) 777-4478; e-mail: osadik@binghamton.edu.

[†] Department of Chemistry, State University of New York at Binghamton.

[‡] Department of Computer Science, State University of New York at Binghamton.

[§] Rensselaer Polytechnic Institute.

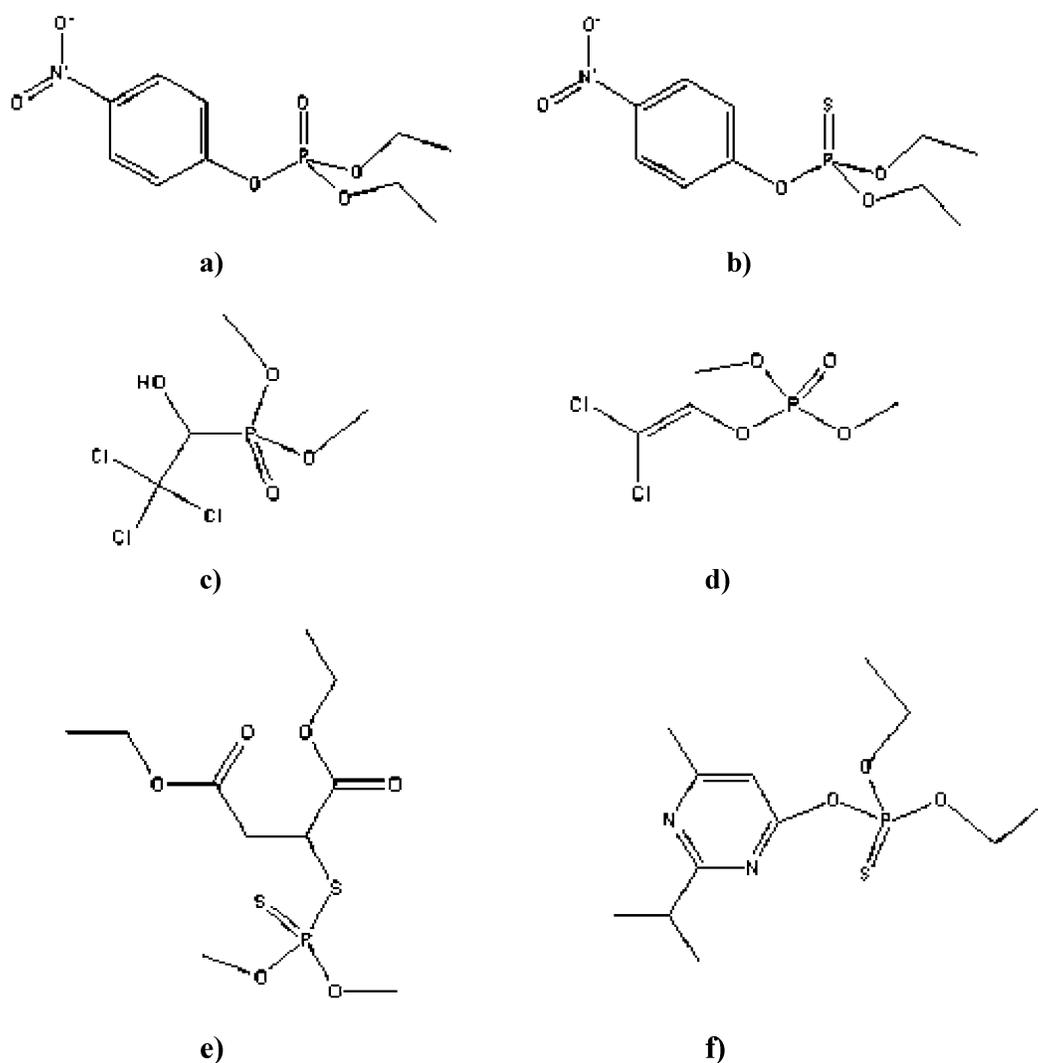


Figure 1. Structures of the organophosphates investigated: (a) paraoxon, (b) parathion, (c) trichlorfon, (d) dichlorvos, (e) malathion, and (f) diazinon.

to be a potential solution to this challenge.¹⁰

In recent years, a family of chemical sensors commonly referred to as the electronic nose (EN) has been widely investigated. A typical EN system consists of an array of chemical sensing elements coupled to headspace sampling, pattern recognition modules, and appropriate transducers.^{11–14} The system employs metal oxides, quartz crystal arrays, surface acoustic wave devices, electrochemical cells, and conducting polymers or a combination of these sensors to mimic the human sense of smell. When used in an array, the sensitivity of an individual sensor is of fundamental importance. The sensor should exhibit high cross-reactivity for the maximum number of components being determined. This requirement is critical for better analytical performance.

Previous attempts to improve the sensitivity of EN systems include the use of conventional pattern recognition techniques,^{11,12,14} increasing the sensitivity of the sensing elements,^{12,14} increasing the amount of volatile compounds reaching the sensor,¹⁵ generating diversity through combinatorial polymer synthesis,^{11,13,14} improving sampling methods,¹⁶ and controlling the effluent flow rate and inadequate temperature control of the effluent-transfer lines.¹⁷ For some types of analytes, sensitivity in the sub-ppm range has been

recorded.^{12–14} However, the scattering of the data obtained can be close to 50%, which dramatically reduces the precision of the measurement.¹⁸ Holberg proposed two methods for countering drift in an electronic sensor array.¹⁹ The first is a self-organizing classifier that stores the patterns of different gas responses. The second is to design the sensor as a dynamic system. Other limitations with these systems are mainly linked to difficulty in calibration, poisoning of the sensing elements, and changes in response time with concentration. Overcoming these limitations requires a greater understanding of the sensor–analyte interactions at the molecular level. Novel intelligent algorithms are urgently needed to process signal patterns in sensor arrays.

Pattern recognition (PR) techniques utilize modern mathematical methods based on multivariate statistics and numerical analysis to elucidate the relationships of multidimensional data sets. These techniques can improve analytical measurements by enhancing the extraction of chemical information from chemical data. They can also reduce the effects of interference and improve selectivity of analytical measurements. Fundamental requirements of PR as applied to sensors include the following: (i) an analyte can be represented as a set of sensor responses; (ii) relationships

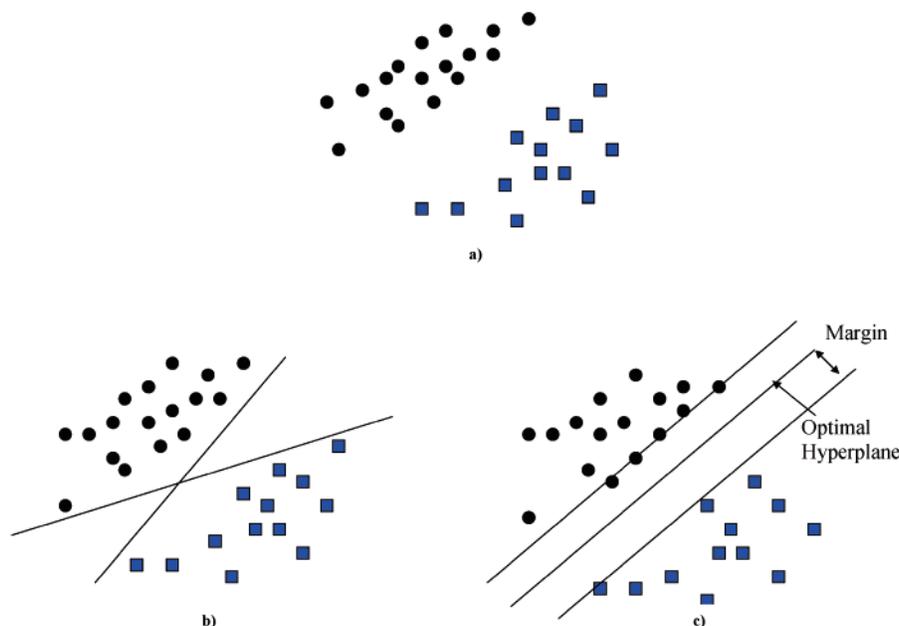


Figure 2. Support vector machines: (a) Hypothetical two-dimensional descriptor space for pairwise analytes/points. (b) Linear separations: There are many hyperplanes that could separate the data. (c) Optimal separating hyperplane: The one that separates the data with maximal margin.

that can be extrapolated to untested analytes from similar classes; and (iii) finding relationships between analytes and their responses that can be tested and verified to a set of tested benchmark analytes. An exhaustive review of computational methods for the analysis of chemical sensor array data from 1994 to 1999 was recently published.¹⁵ The electronic nose information is traditionally obtained by feature extraction using principle component analysis (PCA). PCA, while useful, has several drawbacks because it (i) realizes only input–output mappings and (ii) cannot usually separate independent subsignals from their linear mixture.²⁰

We hereby report the integration of a new strategy to predict, and correctly classify, CWAs using sensor arrays combined with linear support vector machines (SVMs). SVMs are a new and radically different type of classifiers or “learning machines” that use a hypothesis space of linear functions in a high-dimensional feature/space. SVMs are generally trained with learning algorithms originating from optimization theory and that implement a learning bias derived from statistical learning theory. The use of SVMs for computational intelligence is a recent development, and certainly unknown for analytical monitoring of CWAs.

Support Vector Machines. Several texts provide extensive backgrounds to develop the mathematical foundation of support vector machines.^{21–23} In the context of classifying CWAs, specifically organophosphates, the objective of SVMs is to construct an “optimal hyperplane” as the decision surface such that the margin of separation between two different chemical substances is maximized. SVMs are based on the fundamental ideas of (i) structural/empirical risk minimization (SRM/ERM), (ii) the Vapnik–Chervonenkis (VC) dimension, (iii) the constrained optimization problem, and (iv) the SVM decision rule.

The key concepts of SVMs will only be summarized here, including linear support vector machines and the theoretical concept of why SVMs will provide a global minimum, whereas neural networks cannot. In the simplest form, the

support vector machine is a linear classifier. By a linear model we mean a hyperplane that divides the descriptor space into two parts. Consider a set of two different compounds in a certain descriptor space (Figure 2a). With a separating hyperplane, one compound set (circular points) lies, for instance, in one-half of the descriptor space and the other set (square points) in the other half. Points \mathbf{x} on a hyperplane satisfy the equation

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

for some weight vector $\mathbf{w} \in \mathbf{R}^n$ and bias $b \in \mathbf{R}$.

Assume that the circular points lie in the positive half-space and the square points in the negative half-space. It follows that for the circular points

$$\mathbf{w} \cdot \mathbf{x} + b > 0$$

For a given hyperplane, the score of a compound is the signed distance to the hyperplane, which is

$$(\mathbf{w} \cdot \mathbf{x} + b) / (\|\mathbf{w}\|_2)$$

and the distance is the absolute value of this quantity. Therefore, the compounds on the plane have score zero, the compounds in the negative space have a negative score, and the rest have a positive score. In reality, many different hyperplanes may separate the compound correctly (Figure 2b). Support vector machines choose a particular separating hyperplane referred to as the “maximum margin hyperplane” or the “optimal hyperplane” (Figure 2c). The margin of a separating hyperplane is the minimum distance of any labeled data point to the hyperplane. The larger the margin, the clearer the separation between the two sets of compounds. For this reason, the optimal hyperplane is regarded as a robust classifier. Usually only a small set of vectors called support vectors line up closest to the decision boundary; that is, the value of their distance to the boundary equals the margin.

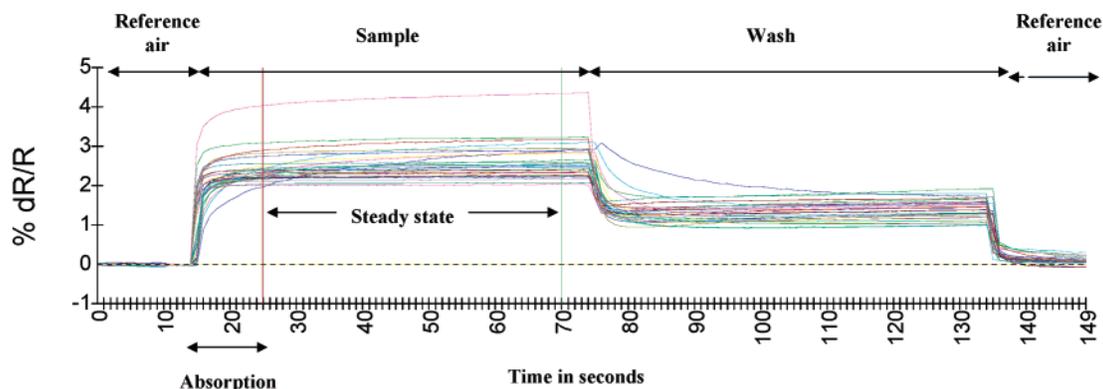


Figure 3. Typical showing of the raw data acquisition process for diazinon. (5 ppm in acetonitrile): reference air was first sampled to produce a baseline response for 15 s; analyte and solvent vapors were then exposed to the 32 sensors for 60 s; the changes in base resistance (dR/R) versus time were recorded for 32 sensor channels; the wash and reference followed for 60 and 15 s, respectively.

Properly designed SVMs should have a good performance on untested data because of their ability to generalize and scale up to more complex problems. The fact that the margin does not depend on input dimensionality means it is immune to the curse of dimensionality. SVMs have been successfully applied to a variety of classification problems including text categorization,²⁴ handwritten digit recognition,^{24–27} gene expression analysis,²⁸ and simple chemical and mixtures recognition.²⁸ In this paper, our aim is to investigate the use of multiarray sensors coupled with support vector machines for the detection of organophosphate nerve agent simulants. This approach reduced the number of false negative errors by 173%, while making no false positive errors when compared to the OsmeTech baseline performance.²⁹

EXPERIMENTAL SECTION

Materials. Malathion, parathion, paraoxon, trichlorfon, diazinon, and dichlorvos were purchased from Accustandard Inc., New Haven, CT, and used as supplied. A 2% aqueous solution of 1-butanol solution was used as the wash solution for the sensor arrays. Gas sensing measurements were conducted using an AS32/8S Labstation and A32/50 series from OsmeTech, CA (Serial Number 32-A11-00-72-058) with accompanying software. The A32S instrument consists of a sample conditioning station (A8S), a sample analyzer, and an A32S computer software package. The analyzer is made up of a detector having an array of 32 conducting polymer sensors with changing electrical resistance when a volatile chemical adheres to them. The change in the electrical resistance of each sensor is monitored and recorded to produce a characteristic pattern or fingerprint.

Procedure. Data Acquisition. 100 microliters each of the organophosphate (OP) pesticides (0.5 ppm in acetonitrile) was introduced into the OsmeTech sample bags (with dimension 7.0×6.5 in.). The reference air was introduced into the bags for 75 s and the bags were equilibrated at ambient temperature for 30 min. Data were collected by passing a sample of the analyte over the sensor array. The changes in resistance versus time for each sensor channel were recorded for all 32 sensors. Reference air was sampled first to produce a baseline response for 15 s. An analyte was then exposed to the conducting polymers sensor arrays for 60 s. Every 1/100th of a second changes in dc resistance of the film is recorded for each sensor in the array. The intensity was calculated as: (dc resistance after absorption – dc

resistance before absorption)/dc resistance before absorption. Also recorded is the time since the start of the experiment, the temperature of the humidity sensor, the in-line temperature of the sample vapor, and the humidity of the sample vapor. After the sample had been exposed to sensor arrays, the sensors were washed with a 2% 1-butanol solution for another 60 s. Finally, the samples were exposed to reference air for 15 s to reproduce the baseline response. The sensor arrays were then exposed to another analyte run and the process is repeated. A total of 250 runs were recorded for each of the six organophosphates. A typical run, showing the data acquisition process for diazinon, is shown in Figure 3.

Preprocessing. Resistance values were digitized using an analogue to digital converter and transferred to a PC for analysis. The type of computer used was a Hewlett-Packard HP Kayak PC Workstation x86 Family 6 Model 5 Stepping 1 with 64884 KB RAM. The raw data file sent to the PC consists of an air baseline, a drift-corrected air baseline, and a set of 35 digitized values for each hundredth of a second that the experiment was running. The 35 values include a separate reading for each of the 32 conducting polymer films showing change in resistance from baseline, temperature, humidity, and temperature of the humidity sensor. Preprocessing of the raw time series data was performed before sending values on to the neural network. In addition to reducing the amount of information that the neural network has to deal with, preprocessing compensates for sensor drift, compresses the transient response of the sensor array, and reduces sample-to-sample variations. However, the existing data reduction program only allows one to gather data from the equilibrium state. We believe that other additional features, such as the rate of absorption, could provide significant clues to distinguishing between sets of compounds. Therefore, a new data reduction software program was written in MATLAB V. 6.1 with the aim to extract additional information, specifically the rate of absorption of the compounds by the sensors.

The MATLAB V. 6.1 preprocessing program is capable of (i) extracting raw and reduced data from the OsmeTech format, (ii) extracting both steady-state and absorption data from each sample using least-squares regression, (iii) removing outliers from the data set, (iv) normalizing the steady state (A32S) for each sensor, (v) creating a single file containing raw time series data created from averaging all

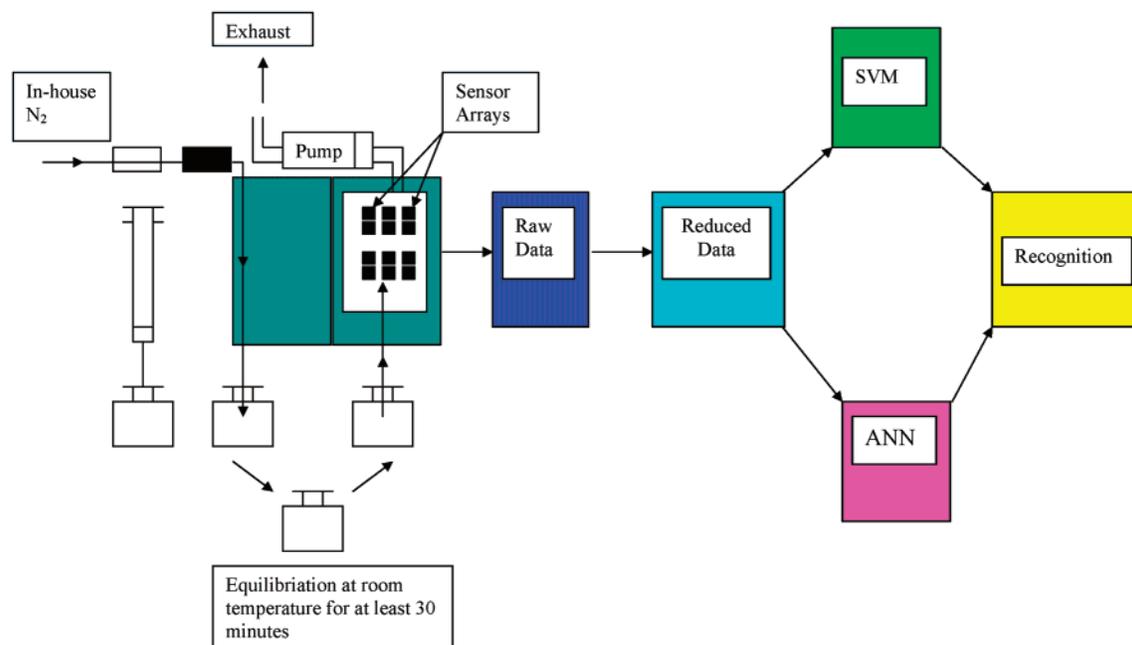


Figure 4. Schematic of the data acquisition, reduction, and processing.

raw time series data in the directory, (vi) assembling reduced data in each subdirectory into one comprehensive data file from which data can be extracted for experiments, and (vii) creating training sets for the SVM experiments by the following: (a) mixing any combination of categories into both a positive and negative case, and randomly sort them, creating an individual file for each reduced feature; (b) normalizing any reduced feature file; (c) split normalizing data into folds for use by the SVM. The data were then processed by the SVM software for “pairwise” and “one versus all” classification. Each SVM run took between 15 and 30 min. A schematic display of the data acquisition, reduction, and processing is shown in Figure 4.

RESULTS AND DISCUSSION

K-fold cross-validation was used to evolve and evaluate the SVMs. This is a statistical process whereby a data set of limited size may be partitioned to simulate a larger data set. This enables a more robust evaluation of the evolved neural network’s generalization ability and expected performance. Initial experiments were performed using a 5-fold cross-validation process, where 200 samples were used in the training set and 50 samples used in the validation set. The procedure allowed the network to be trained on the most possible data, while reducing the likelihood that the partitioning of the data had introduced a bias into the results (as could happen if the “difficult” cases were all in the training set and the validation set contained all “easy” cases). Several performance measures were used to evaluate the conventional artificial neural network (ANN) as well as new machine learning techniques.

Standard three layer artificial neural networks, trained by back-propagation, and using the sigmoid activation function, are often criticized as “black boxes”, which are almost always trained to a local minimum, with the resultant degradation of performance. This fact may be theoretically demonstrated by the following proof. Assume there exist *N* observations from an organophosphate data set. Each observation (or

training example) consists of a vector \mathbf{x}_i containing the input pattern and a corresponding known classification y_i . The objective of the learning machine would be to formulate a mapping $\mathbf{x}_i \rightarrow y_i$. Now consider a set of functions $f(\mathbf{x}, \alpha)$ with adjustable parameters α that define a set of possible mappings $\mathbf{x} \rightarrow f(\mathbf{x}, \alpha)$. Here, \mathbf{x} is given and α is chosen. In the case of a traditional neural network of fixed architecture, the α values would correspond to the weights and biases. The quantity $R(\alpha)$, known as the *expected* (or *true*) risk, associated with learning machines is defined as

$$R(\alpha) = \int (1/2) |y - f(\mathbf{x}, \alpha)| p(\mathbf{x}, y) \, d\mathbf{x} \, dy$$

where $p(\mathbf{x}, y)$ is an unknown probability density function from which the examples were drawn. This risk function is the expected (or true) value of the test (or validation) error for a trained learning machine. It may be shown that the best possible generalization ability of a learning machine is achieved by minimizing $R(\alpha)$, the expected (or true) risk. This generalization bound, for binary classification, holds with the probability of at least $1 - \eta$ ($0 \leq \eta \leq 1$) for all approximating functions that minimize the expected (or true) risk.

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \sqrt{\left(\frac{h \left(\log \left(\frac{2N}{h} \right) + 1 \right) - \log \left(\frac{\eta}{4} \right)}{N} \right)}$$

The first term on the right-hand side of the above expression is known as the “empirical risk”. The empirical risk $R_{\text{emp}}(\alpha)$ is defined as

$$R_{\text{emp}}(\alpha) = \frac{1}{2N} \sum_{i=1}^N |y_i - f(\mathbf{x}_i, \alpha)|$$

This function is a measure of the error rate for the training set for a fixed, finite number of observations. This value is fixed for a particular choice of α and a given training set

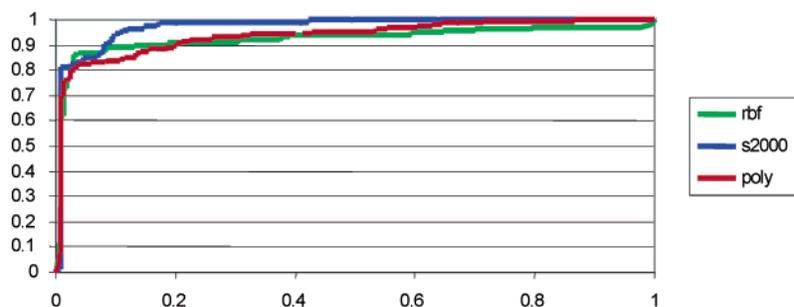


Figure 5. ROC (receiver operating curves) for parathion vs dichlorvos.

(\mathbf{x}_i, y_i) . The second term in the true risk expression is the “Vapnik–Chervonenkis (VC) confidence interval.” This term is a function of the number of training samples N , the probability value η , and the VC dimension h . The VC dimension is the *maximum* number of training samples that can be learned by a learning machine without error for *all* possible labeling of the classification functions $f(\mathbf{x}, \alpha)$ and is, therefore, a measure of the capacity of the learning machine. In traditional neural-network implementations, the confidence interval is fixed by choosing a network architecture *a priori*. The risk function is typically minimized by minimizing the empirical risk through adjustment of weights and biases, resulting in a local minimum. Consequently, neural networks are trained based on the empirical risk minimization (ERM) principle. In a SVM design and implementation, not only is the empirical risk minimized but also the VC confidence interval by using the principles of structural risk minimization (SRM). Therefore, SVM implementations simultaneously minimize the empirical risk as well as the risk associated with the VC confidence interval, as defined in the true risk expression above. The true risk expression above also shows that as $N \rightarrow \infty$, the empirical risk approaches the true risk because the VC confidence interval risk approaches zero. The reader may recall that obtaining larger and larger sets of valid training data would sometimes produce (with a great deal of training experience) a better performing neural network (NN), which resulted from classical training methods. This restriction is not incumbent on the SRM principle and is the fundamental difference between training NNs and training SVMs.

Several SVM kernels are utilized in this analysis. These are the

dot product kernel: $K(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i \cdot \mathbf{x})$

GRBF kernel: $K(\vec{x}_i, \vec{x}) = e^{-\|\vec{x}_i - \vec{x}\|^2 / 2\sigma^2}$

polynomial kernel: $K(\vec{x}_i, \vec{x}) = (\vec{x}_i \cdot \vec{x} + 1)^d$

S2000 kernel: $\|X - Y\|^2$

Although this optimization initially focused on the ROC A_z area achieved under 5-fold cross-validation, the authors were particularly interested in achieving and measuring the best possible performance at higher sensitivity. The ROC curve is a function of the false alarm rate (FAR = 1-specificity), plotted along the abscissa, vs the sensitivity, which is plotted along the ordinate. The area under the curve is called the A_z index. The closer to unity, the better the A_z value. For example, an A_z value of 0.5 represents chance

performance, 0.7 is okay, 0.8 is good, and 0.9 or greater is excellent. A typical ROC curve realized during the investigation is shown in Figure 5. Measures of performance (MOP) parameters associated with ROC curve analysis are as follows:

(a) Sensitivity $\{TP/(TP + FN)\}$, which is defined as the likelihood that an event will be detected if that event is present (where TP is a true positive and FN is a false negative event). Sensitivities of 100%, 98%, and 95% represent false negative errors of 0%, 2%, and 5%, respectively. The objective is to have the system operate at 100% sensitivity.

(b) Specificity $\{TN/(TN + FP)\}$, which is the likelihood that the absence of an event will be detected, given that the event is absent (where TN is a true negative and FP is a false positive event). In a like manner, specificities of 100%, 98%, and 95% represent false positive errors of 0%, 2%, and 5%. The objective, then, is to design the learning system so that we increase the specificity at 100% sensitivity (making no false negative errors). By increasing the specificity, we reduce the number of false positive errors.

Another point should be made here. For SV learning machines trained to a global minimum, the A_z index will generally be in the region of 0.97–0.99, which is outstanding performance. When these MOP’s are achieved with “small” databases (<500 samples), little, if any, difference will be observed in the specificities and PPVs for sensitivities of 95%, 98%, and 100%. This situation exists because of computational “round off” error because of the small differences in false positives and negatives at the threshold settings corresponding to these specificities and sensitivities.

(c) Positive predictive value (PPV) $\{TP/(TP + FP)\}$, which measures the likelihood that a signal of an event is associated with that event, given that such a signal occurred.

In addition to the A_z index, which is the area under the ROC curve, this work measures the ability of the SVMs to increase the specificity and PPV at the relevant higher sensitivities. That is, at 100% sensitivity the SVM models make no type II errors (positive cases identified as negatives). Consequently, increasing the specificity percentage at a relevant 100% sensitivity is a direct measure of decreasing the type I errors (negative cases identified as positives), while the system makes no type II errors. In a like manner, increasing the specificity percentage at 98% sensitivity (missing 2% of the type II errors) will then quantitatively describe learning machine performance in decreasing the type I errors. Similarly, increasing the PPV at the higher sensitivities provides a similar, but somewhat different, performance measure. In the following sections we study the accuracy of the SVM classifiers in identifying structurally similar organophosphate nerve agents such as paraoxon and parathion.

Table 1. Comparison of Measures of Performance of ANN and S2000 Kernel for the Parathion vs Dichlorvos Pair

	ANN	S2000	% improvement
A_z	0.8021	0.9853	23
specificity, 100%	0.3280	0.8963	173
PPV, 100%	0.6460	0.9027	40
specificity, 98%	0.3280	0.8963	173
PPV, 98%	0.6460	0.9027	40

Table 2. Comparison of Measures of Performance of ANN and RBF Kernel for the Parathion vs Dichlorvos Pair

	ANN	RBF	% improvement
A_z	0.8021	0.944	18
specificity, 100%	0.3280	0.683	108
PPV, 100%	0.6460	0.868	34
specificity, 98%	0.3280	0.683	143
PPV, 98%	0.6460	0.868	37

Table 3. Comparison of Measures of Performance of ANN and the Polynomial Kernel for the Parathion vs Dichlorvos Pair

	ANN	polynomial	% improvement
A_z	0.8021	0.9759	22
specificity, 100%	0.3280	0.7977	143
PPV, 100%	0.6460	0.8834	37
specificity, 98%	0.3280	0.7977	143
PPV, 98%	0.6460	0.8834	37

The performance of the three different kernels is evaluated (i.e., the RBF (radial basis function) kernel, the S2000 kernel, and the polynomial kernel).

Parathion and Dichlorvos. The performance of the ANN was compared to SVMs with the S2000, Gaussian, and polynomial kernels. Tables 1, 2, and 3 show the performance of the S2000, RBF, and polynomial kernels, respectively, in comparison with the ANN data obtained for parathion and dichlorvos pair. The tables show that there was a 23% improvement in the ROC A_z index by using S2000 (Table 1), 18% improvement using the RBF kernel (Table 2), and 22% improvement using the polynomial kernel (Table 3). The ROC curves for the parathion vs dichlorvos pair resulting from the three kernels is shown in Figure 5. Our results also show that there was a significant improvement (173%) of specificities at 100% and 98% sensitivities on using the S2000 kernel (Table 1). This means that the number of false negative errors was reduced by 173%. Specificities at 100% and 98% for the other kernels were also very impressive as they registered an improvement of 108% and 143%, respectively, when compared to the ANN baseline (Tables 2 and 3). PPV at 100% and 98% sensitivities recorded a 40% improvement on using the S2000 (Table 1), 34% improvement on using the RBF (Table 2), and 37% improvement on using the polynomial kernel (Table 3). On the basis of these three MOPs, we can conclude that the three kernels studied are significantly better than the ANN software in the AromaScan for the parathion and dichlorvos base pair. We can also conclude that, of the three kernels studies, the S2000 exhibited the best performance.

Parathion vs Paraoxon. The comparison between the parathion and paraoxon pair should be very interesting because of their very close structural similarities. As previously stated, the only difference in the structures of the two compounds is that the P=O bond in paraoxon is replaced by a P=S bond in parathion. The performance of the

Table 4. Comparison of Measures of Performance of the Three Kernels for the Parathion vs Paraoxon Pair (% Improvements over the ANN Are Shown in Brackets)

	RBF	S2000	polynomial
A_z	0.9270 (16%)	0.840 (5%)	0.9910 (24%)
100% specificity	0.7633 (133%)	0.8701 (165%)	0.8739 (166%)
100% PPV	0.7304 (13%)	0.8359 (29%)	0.8366 (30%)
98% specificity	0.7633 (133%)	0.8701 (165)	0.8739 (166%)
98% PPV	0.7304 (13%)	0.8359 (29%)	0.8366 (30%)

Table 5. Comparison of Measures of Performance of the Three Kernels for the Dichlorvos vs Trichlorfon Pair (% Improvements over the ANN Are Shown in Brackets)

	RBF	S2000	polynomial
A_z	0.9982 (24%)	0.9988 (25%)	0.9999 (25%)
100% specificity	0.9934 (203%)	0.9922 (203%)	0.9965 (204%)
100% PPV	0.9934 (54%)	0.9942 (54%)	0.9969 (54%)
98% specificity	0.9934 (203%)	0.9965 (204%)	0.9965 (204%)
98% PPV	0.9933 (54%)	0.9968 (54%)	0.9968 (54%)

polynomial kernel was the best for this particular pair because it exhibited the greatest improvement in all the three MOPs parameters measured when compared to the baseline ANN (Table 4). The ROC A_z , specificities, and PPV for this kernel improved by 24%, 166%, and 30%, respectively. The corresponding values for S2000 and RBF kernels were 5%, 165%, and 29% and 16%, 133%, and 13%, respectively. These are excellent results considering only one-atom differences in these nerve agent simulants.

Trichlorfon vs Dichlorvos. The comparison between this pair undoubtedly gave the best results. The ROC A_z values of 0.9982, 0.9988, and 0.9999 for the RBF, S2000, and polynomial kernels represent improvements of between 24 and 25% over the ANN baseline (Table 5). The specificities and PPVs also exhibited near perfect values with improvements of 54% and between 203 and 204%, respectively, over the ANN baseline. In summary, all three SVMs are essentially perfect classifiers for this chemical pair.

Detection and Classification. We tested our new database to positively identify unknown samples as one of "m" chemicals in their database. This is a multiclass problem in which the output domain is changed from $Y = \{-1, 1\}$ to $Y = \{1, 2, 3...m\}$. Although SVMs are mainly used for two-class problems, they can be extended to handle multiclass problems using a voting scheme to combine the outputs of several binary SVMs trained on different pairs of chemicals. The most successful voting scheme of those we have tried yet was pairwise classification in which the outputs from $K(K-1)/2$ binary SVMs are used to fill a square $K \times K$ table. SVM (a, b) and SVM (b, a) have reflectional symmetry in the zero plane, so each of the $K(K-1)/2$ classifiers fill two entries ($[a, b]$, $[b, a]$) in the table. A binary classifier decides whether a point x belongs to class a or b . The probability that x belongs to class a , given that x is in either class a or b , can be written as

$$P_{ab} = P(xEa|x, xEa \cup b)$$

With P_{ab} , we can calculate the estimate Pa of the a posteriori probability $P(xEa|x)$ by using a $K \times K$ table of P_{ab} (the chemical voted for) and $P_{ba} = 1 - P_{ab}$ (the chemical not voted for) as

$$P_i = 2/K(K-1)^* \sum P_{ab} \quad (2)$$

Table 6. S2000 Classification Results: True vs Predicted (Row vs Columns)

	diazinon	malathion	parathion	trichlorfon	dichlorvos	paraoxon
diazinon	20	0	0	0	0	0
malathion	0	19	0	0	1	0
parathion	0	0	20	0	0	0
trichlorfon	0	0	0	20	0	0
dichlorvos	0	0	0	0	20	0
paraoxon	0	0	0	0	0	20

Table 7. ANN Classification Results: True vs Predicted (Row vs Columns)

	diazinon	malathion	parathion	trichlorfon	dichlorvos	paraoxon	unknown
diazinon	7	0	11	0	0	0	2
malathion	0	17	0	0	0	3	0
parathion	0	0	20	0	0	0	0
trichlorfon	16	0	2	0	0	1	1
dichlorvos	0	0	3	0	17	0	0
paraoxon	1	0	0	14	5	0	0

The SVM decision output $Fab(x)$ is not a probability value, so we have to

1. Normalize $Fab(x)$ such that the output is ± 1 .

2. Map the SVM output to $Pab = Fab(x) + 0.5$. In a binary classifier, a vote of 100% certainty for a chemical is either “-1” (100% *a*, 0% *b*) or “+1” (0% *a*, 100% *b*). A vote of 0 would indicate absolutely no inclination toward one chemical or the other (50% *a*, 50% *b*).

The $K \times K$ table is thus filled with values of Pab , which can in turn be input into a class decision function. The classifying decision is made by adding the $(K - 1)$ votes for each chemical together (add up the rows) and choosing the chemical with the highest vote. In this scheme the strength of a classification for one chemical automatically weakens the chances for the other chemical to win. If there is no strong winner (no vote is higher than a given threshold), it indicates that no chemical was favored by all SVMs trained on it. Previously, we had created a pairwise classifier to detect one of the four organophosphates: dichlorvos, trichlorfon, paraoxon, and malathion. Six binary support vector machines were trained including the following: (1) dichlorvos vs trichlorfon; (2) dichlorvos vs paraoxon; (3) dichlorvos vs malathion; (4) trichlorfon vs paraoxon; (5) trichlorfon vs malathion, and (6) paraoxon vs malathion. The training file for each classifier contained the following features for each sample:

- (1) The steady-state feature of 32 sensors (Figure 3).
- (2) The absorption rate feature of 32 polypropylene sensors (Figure 3).
- (3) Two temperature values (steady state and absorption rate).
- (4) Two humidity values (steady state and absorption rate).
- (5) Two temperature and humidity sensor values (steady state and absorption rate).

The results of this classifier were nearly perfect (Table 6) Dichlorvos, trichlorfon, and paraoxon were perfectly predicted. Only 1 run out of 20 runs for malathion was misclassified and was classified as dichlorvos. These results are very good when compared to the results realized by the ANN (Table 7) only 1 sample out of 80 samples tried was mispredicted. Clearly, the training data had sufficient information to make good classification decisions. On the other hand, the ANN had only one perfect prediction with parathion. The other six organophosphates exhibited various levels of mispredictions. For example, trichlorfon and

paraoxon were totally mispredicted. The prediction successes of both malathion and dichlorvos was 85% while that of diazinon was 35%. Our future study will attempt to train the SVM kernels using different concentrations of one analyte so as to enable the quantification of concentrations.

CONCLUSION

We have integrated multiarray sensors with support vector machines to predict and correctly classify organophosphate agent simulants. We designed and evaluated a new SVM protocol for more accurate classification software. Analytical sensor signals generated were fed into the SVM software and these were used to generate a database of over 500 proven analyte samples. The best performing SVM was designated as having the most accurate specificity. The results of this research have demonstrated that all the three kernels studied showed significant improvements in all the MOPs studied when compared to the ANN baseline for all the organophosphate pairs investigated. Our results showed that in all the three MOPs the three kernels recorded significant improvements when compared with ANN. In all the pairs investigated, the ROC A_z values, specificities, and PPVs increased by between 5% and 25%, 108% and 204%, and 13% and 54%, respectively, when compared to the ANN baseline measures of performance. Despite having close structural similarity, the MOP values for the parathion and paraoxon pair showed significant improvement when compared to the ANN baseline. Using the S2000 kernel, dichlorvos, trichlorfon, and paraoxon were perfectly predicted. Only 1 run out of 20 runs for malathion was mispredicted. Further experiments are going on to investigate the possibility of training the SVM kernels for quantitation of structurally similar analytes at different concentrations.

ACKNOWLEDGMENT

The authors thank the National Science Foundation for financial support of this work (CHE-0210968 and IIS-9979860).

REFERENCES AND NOTES

- (1) (a) Donarski, W. J.; Dumas, D. P.; Heitmeyer, D. P.; Lewis, V. E.; Raushel, F. M. Structure-Activity Relationships in the Hydrolysis of Substrates by the Phosphotriesterase from *Pseudomonas diminuta*. *Biochemistry* **1989**, *28*, 4650–4655. (b) Chapalamadugu, S.; Chaudhry,

- G. S. Microbiological and Biotechnological Aspects of Metabolism of Carbamates and Organophosphates. *Crit. Rev. Biotechnol.* **1992**, *12*, 357–389. (c) Food and Agricultural Organization of the United Nations (FAO), Rome, *FAO Prod. Yearbook* **1989**, *43*, 320.
- (2) (a) United States Department of Agriculture. *Agricultural Statistics*; United States Government Printing Office: Washington, DC, 1992; p 395. (b) Compton, J. A. *Military Chemical and Biological Agents*; Telford Press: Caldwell, NJ, 1988; p 135.
- (3) Sherma, J. Pesticides. *Anal. Chem.* **1993**, *65*, R40–R54.
- (4) Sherma, J. Pesticides. *Anal. Chem.* **1995**, *67*, R1–R20.
- (5) (a) Everett, W. R.; Rechnitz, G. A. Mediated Bioelectrocatalytic Determination of Organophosphorus Pesticides with a Tyrosinase-Based Oxygen Biosensor. *Anal. Chem.* **1998**, *70*, 807. (b) Mulchandani, A.; Kaneva, I.; Chen, W. Biosensor for Direct Determination of Organophosphate Nerve Agents Using Recombinant *Escherichia coli* with Surface-Expressed Organophosphorus Hydrolase. 2. Fiber-Optic. *Anal. Chem.* **1998**, *70*, 5042. (c) Abad, J. M.; Pariente, F.; Hernandez, Abruna, H. D.; Lorenzo, E. Determination of Organophosphorus and Carbamate Pesticides Using a Piezoelectric Biosensor. *Anal. Chem.* **1998**, *70*, 2848.
- (6) (a) Mulchandani, A.; Mulchandani, A.; Chen, W. Amperometric Thick-Film Strip Electrodes for Monitoring Organophosphate Nerve Agents Based on Immobilized Organophosphorus Hydrolase. *Anal. Chem.* **1999**, *71*, 2246. (b) Mulchandani, P.; Chen, W.; Mulchandani, A. Flow Injection Amperometric Enzyme Biosensor for Direct Determination of Organophosphate Nerve Agents. *Environ. Sci. Technol.* **2001**, *35*, 2562. (c) Mulchandani, P.; Chen, W.; Mulchandani, A.; Wang J.; Chen, L. Amperometric microbial biosensor for direct determination of organophosphate pesticides using recombinant microorganism with surface expressed organophosphorus hydrolase. *Biosens. Bioelectron.* **2001**, *16*, 433.
- (7) Karousos, N. G.; Aouabdi, S.; Way, A. S.; Reddy, S. M. Quartz crystal microbalance determination of organophosphorus and carbamate pesticides. *Anal. Chim. Acta* **2002**, *469*, 189.
- (8) Wang, J.; Krause, R.; Block, K.; Musameh, M.; Mulchandani, A.; Mulchandani, P.; Chen, W.; Schoning, M. J. Dual amperometric-potentiometric biosensor detection system for monitoring organophosphorus neurotoxins. *Anal. Chim. Acta* **2002**, *469*, 197–203.
- (9) (a) Sadik O. A.; Land W.; Wang J. Targeting Chemical and Biological Warfare Agents at the Molecular Level. *Electroanalysis* **2003**, *15*, 1149 (b) Ciucu A. A.; Negulescu, C.; Baldwin, R. P. Detection of pesticides using an amperometric biosensor based on ferophthalocyanine chemically modified carbon paste electrode and immobilized bienzymatic system. *Biosens. Bioelectron.* **2003**, *18*, 303. (c) Sadik, O. A.; Van Emon J. M. Applications of electrochemical immunosensors to environmental monitoring. *Biosens. Bioelectron.* **1996**, *11*, 9. (d) Sadik, O. A.; Ngundi, M. M.; Yan, F. *Biotechnology and Bioprocess Engineering* **2000**, *5*, 407.
- (10) Grate J. W.; RosePehrsson S.; Venezky, D. L.; Klusty, M.; Wohltjen, H. Smart sensor system for trace organophosphorus and organosulfur vapor detection employing a temperature-controlled array of surface acoustic wave sensors, automated sample preconcentration, and pattern recognition. *Anal. Chem.* **1993**, *65*, 1868–1881.
- (11) (a) Grate J. W.; Abraham M. H. Solubility interactions and the design of chemically selective sorbent coatings for chemical sensors and arrays. *Sens. Actuators, B* **1991**, *3*, 85. (b) Steter, J. R.; Jurs, P. C.; Rose, S. L. *Anal. Chem.* **1986**, *58*, 860. (c) Pearce T. C.; Gardner J. W.; Friel, S.; Bartlett, P. N.; Blair, N. Electronic Nose for Monitoring the Flavor of Beers. *Analyst* **1993**, *118*, 371. (d) Rose-Pehrsson, S. L.; Grate, J. W.; Ballantine, D. S. Jr.; Jurs, P. C. Detection of Hazardous Vapors Including Mixtures Using Pattern-Recognition Analysis of Responses from Surface Acoustic-Wave Devices. *Anal. Chem.* **1988**, *60*, 2801. (e) Zellers E. T.; Batterman, S. A.; Han, M.; Patrash, S. J. Optimal Coating Selection for the Analysis of Organic Vapor Mixtures with Polymers-Coated Surface-Acoustic-Wave Sensor Arrays. *Anal. Chem.* **1995**, *67*, 1092. (f) Nagle, H. T.; Schiffman, S. S.; Gutierrez-Osuna, R. Electronic nose technology: an overview. *IEEE Spectrum* **1998**, Sept, 22. (g) Lonergan, M. C.; Severin, E. J.; Dolema, B. J.; Beaber, S. A.; Grubbs, R. H.; Lewis, N. S. Array-Based Vapor Sensing Using Chemically Sensitive, Carbon Black-Polymer Resistors. *Chem. Mater.* **1996**, *8*, 2298. (g) Grate, J.; Abraham, M. Solubility interactions and the design of chemically selective sorbent coatings for chemical sensors and arrays. *Sens. Actuators* **1991**, *3*, 85.
- (12) (a) Kress-Rogers, E., Ed. *Handbook of Biosensors and Electronic Noses*; CRC Press: Boca Raton, FL, 1997. (b) Persaud, K. C.; Khaffaf, S. M.; Hobbs, P. J.; Sneath, R. W. Assessment of conducting polymer odor sensors for agricultural malodor measurements. *Chem. Senses* **1996**, *21*, 495. (c) Masila, M. M.; Sadik, O. A. Electronic Nose for Detection of Polyaromatic Hydrocarbons. *ACS Symp. Ser.* **2000**, *762*, 37.
- (13) (a) Persaud, K. C.; Khaffaf, S. M.; Payne, J. S.; Pisanelli, A. M.; Lee, D.; Byun, H. Sensor array techniques for mimicking the mammalian olfactory system. *Sens. Actuators, B* **1996**, *35*, 267. (b) Persaud, K. C.; Khaffaf, S. M.; Hobbs, P. J.; Sneath, R. W. Assessment of conducting polymer odor sensors for agricultural malodor measurements. *Chem. Senses* **1996**, *21*, 495.
- (14) (a) Walt, D. R.; Dickinson, T. A.; White, J.; Kauer, J.; Johnson, S.; Engelhart, H.; Sutter, H.; Jurs, P. Optical sensor arrays for odor recognition. *Biosens. Bioelectron.* **1998**, *13*, 697. (b) Dickinson, T. A.; White, J.; Kauer, J.; Walt, D. R. A chemical-detecting system based on a cross-reactive optical sensor array. *Nature* **1996**, *382*, 697. (c) Rakow, N. A.; Suslick, K. S. A colorimetric sensor array for odor visualization. *Nature* **2000**, *406*, 710.
- (15) (a) Jurs, P. C.; Bakken, G. A.; McClelland, H. E. Computational Methods for the Analysis of Chemical Sensor Array Data from Volatile Analytes. *Chem. Rev.* **2000**, *100*, 2649. (b) Hines, E. L.; Lobet, E.; Gardner, J. W. Electronic noses: A review of signal processing techniques. *IEEE Proc. Circuits Devices Syst.* **1999**, *146*, 297. (c) Gardner, J. W.; Batlett P. N. Pattern Recognition in Gas Sensing. In *Techniques and Mechanisms in Gas Sensing*; Moseley, P. T., Norris, J., William, D., Eds.; Adam-Hilger: Bristol, UK, 1981.
- (16) Masila, M.; Breimer, M.; Sadik, O. A. *Electronic Noses and Sensor Array Based Systems – Design and Applications*; Hurst, W. J., Ed.; Technomic: Lancaster, PA, 1999; p 27.
- (17) Roca, S.; Delgado, I.; Ferrer Correia, A. J.; Barroa, A.; Wells, P. Application of an Electronic Aroma Sensing System to Cork Stopper Quality Control. *J. Agric. Food Chem.* **1998**, *46*, 145.
- (18) (a) Stuetz, R. M.; Fenner, R. A.; Engin, G. Assessment of odours from sewage treatment works by an electronic nose, H₂S analysis and olfactometry. *Water Res.* **1999**, *33*, 453. (b) Stuetz, R. M.; Fenner, R. A.; Engin, G. Characterization of wastewater using an electronic nose. *Water Res.* **1999**, *33*, 442.
- (19) Gibson, T. D.; Presser, O.; Holberg, J. N.; Marshall, R. W.; Corcoran, P.; Lowery, P.; Ruck-Keene, E. A.; Heron, S. Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose. *Sens. Actuators, B* **1997**, *44*, 413.
- (20) Pal, R. N. Soft computing for feature analysis. *Fuzzy Sets Syst.* **1999**, *103*, 201
- (21) (a) Vapnik V. *Statistical Learning Theory*; Wiley: New York, 1998. (b) Vapnik, V. N. *The Nature of Statistical Learning Theory*; Wiley: New York, 1995. (c) Cristianini, N.; Shawne-Taylor, C. *Support Vector Machines and other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, 2000. (d) Platt, J. C. Fast training of Support Vector Machines Using Sequential Minimal Optimization. In *Advances in Kernel Methods: Support Vector Training*; Scholkopf, B., Burges, C., Smola, A., Eds.; MIT Press: Cambridge, MA, 1998; p 185.
- (22) (a) Embrechts, M. J.; Bennet, K. P.; Breneman, C. www.drugmining.com, 2000. (b) Keerthi, S. S.; Shevade, S. K. SMO algorithm for least-squares SVM formulations. *Neural Comput.* **2003**, *15*, 487–507. (c) Evgeniou, T.; Pontil, M.; Poggio, T. Statistical Learning Theory: A primer. *Int. J. Comput. Vision* **2000**, *38*, 9–13. (d) Evgeniou, T.; Pontil, M.; Poggio, T. Regularization Networks and Support Vector Machines. In *Advances in Large Margin Classifiers*; MIT Press: Cambridge, MA, 2000.
- (23) (a) Fung, G.; Mangasarian, O. L. Proximal Support Vector Machine Classifiers. In *Proceedings KDD 2001*, San Francisco, CA, 2001. (b) Poggio, T.; Smale S. The Mathematics of Learning: Dealing with Data. To appear in *Not. AMS*, 2003. (c) Scholkopf, B.; Smola, A. J. *Learning with Kernels*; MIT Press: Cambridge, MA, 2002.
- (24) Suykens, J. A. K.; VandeWalle, J. Least-squares support vector classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300.
- (25) Drucker H.; Wu D.; Vapnik V. N. Support Vector Machine for spam categorization. *IEEC Trans. Neural Network* **1999**, *10*, 1048–1054.
- (26) DeCoste B.; Scholkopf B. Training invariant support vector machines. *Mach. Learn.* **2002**, *46*, 161–190.
- (27) Furey T. S.; Cristianini N.; Duffy N.; Bednarski D. W.; Schummer M.; Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **2000**, *16*, 906–914.
- (28) Distante, C.; Ancona, N.; Siciliano, P. Support vector machines for olfactory signals recognition. *Sens. Actuators, B* **2003**, *88*, 30–39.
- (29) Land, W.; Sadik, O.; Leibensperger, D.; Breimer, M. *Smart Engineering System Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Complex Systems and Artificial Life*, Proceedings of the Artificial Neural Networks in Engineering Conference (ANNIE 2002); Cihan, A., Dagli, H., Buczak, A. L., Ghosh, J., Embrechts, M. J., Ersoy, O., Kernel, S. L., Eds.; ASME Press: New York, 2002; Vol 12, pp 811–816.