Artificial Odour Discrimination System using electronic nose and neural networks for the identification of urinary tract infection.

Vassilis S. Kodogiannis[1]
John N. Lygouras[2]
Andrzej Tarczynski[3]
Hardial S. Chowdrey[4]

[1]School of Informatics, University of Westminster
[2]Dept of Electrical & Computer Engineering, Democritus University of Thrace
[3]Harrow School of Computer Science, University of Westminster
[4]School of Biosciences, University of Westminster

Artificial Odor Discrimination System Using Electronic Nose and Neural Networks for the Identification of Urinary Tract Infection

Vassilis S. Kodogiannis, *Member, IEEE*, John N. Lygouras, Andrzej Tarczynski, *Member, IEEE*, and Hardial S. Chowdrey

***Abstract*—Current clinical diagnostics are based on biochemical, immunological, or microbiological methods. However, these methods are operator dependent, time-consuming, expensive, and require special skills, and are therefore, not suitable for point-of-care testing. Recent developments in gas-sensing technology and pattern recognition methods make electronic nose technology an interesting alternative for medical point-of-care devices. An electronic nose has been used to detect urinary tract infection from 45 suspected cases that were sent for analysis in a U.K. Public Health Registry. These samples were analyzed by incubation in a volatile generation test tube system for 4–5 h. Two issues are being addressed, including the implementation of an advanced neural network, based on a modified expectation maximization scheme that incorporates a dynamic structure methodology and the concept of a fusion of multiple classifiers dedicated to specific feature parameters. This study has shown the potential for early detection of microbial contaminants in urine samples using electronic nose technology.**

***Index Terms*—Electronic nose, microbial analysis, multiple classifiers, neural networks (NNs).**

I. Introduction

THE INCREASED knowledge in bionics and artificial intelligence has revolutionized many areas of human activity. The employment of these approaches in medicine will be no exception. New socioeconomic factors and general globalization of the world requires the development and application of new intelligent diagnostic and therapeutic near-patient or home-based devices to control disease more effectively [1].

Microbial diseases constitute the major cause of death in many developing countries. Bacterial detection requires analytical methods that can satisfy a series of criteria such as short detection time ($t \leq 3$ h), sensitivity (detection of bacterial concentrations during infection $>10^5$ cells/ml) specificity (species accurate identification), and low cost.

Manuscript received September 1, 2006; revised July 29, 2007 and November 16, 2007. First published May 20, 2008; current version published November 5, 2008.

V. S. Kodogiannis is with the Centre for Systems Analysis, School of Computer Science, University of Westminster, London HA1 3TP, U.K. (e-mail: kodogiv@wmin.ac.uk).

J. N. Lygouras is with the Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi 67100, Greece.

A. Tarczynski is with the Centre for Systems Analysis, School of Informatics, University of Westminster, London W1W 6UW, U.K.

H. S. Chowdrey is with the Department of Biomedical Sciences, School of Biosciences, University of Westminster, London W1W 6UW, U.K.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITB.2008.917928

Over the past years, there have been an increasing number of attempts to apply "olfactory" diagnostics based on volatiles production in clinical practice. Microbial–host interaction of certain metabolic disorders can cause early liberation of gas molecules in breath, urine, feces, and sputum or sweat in clinical samples. Even though the introduction of gas chromatographic (GC) techniques and later mass spectrometry (MS) enabled the comprehensive study of possible disease markers, it never evolved to a fully operational and accredited diagnostic tool. Increased capital cost, laborious and time-consuming methods, the severe complexity of human volatile samples, and the need for highly skilled personnel are some of the limitations that emerged through the years [2]. However, the accumulated knowledge helped understanding of the way the human body responds to disease and the role of organic volatile compounds (VOCs) as target markers for future smart diagnostics.

The name "electronic nose" (e-nose) comes from a certain parallel of the measurement concept of the instrument and that of the mammalian olfactory system. e-Noses base the analysis on the cross-reactivity of an array of semiselective sensors. Often the sensitivity of electronic noses is similar to that of human noses but humans are specially gifted in sensing specific compounds. The biological sensitivity can go down to ppt levels with a response time in the order of milliseconds whereas instruments barely go under ppb levels with a response time in the order of seconds (Table I) [3].

In industry, odor assessment is usually performed by human sensory analysis or by GC/MS. The latter technique gives information about volatile organic compounds but the correlation between analytical results and global odor perception is not direct due to potential interactions between several odorous components. The ideal sensors for integration in an e-nose should fulfill the following criteria: high sensitivity; they must respond to different compounds present in the headspace of the sample; high stability and reproducibility; short recovery time; easy calibration; they must also be robust and portable [4].

Various sensor technologies are employed in e-noses. The most popular ones that are now used in commercial instruments are semiconducting metal oxides and electronically conducting polymers. All types of sensors share a common basic principle. The interaction of VOCs with the sensor surface leads to a change of physical properties (conductivity, resistance, frequency) of the sensor, which is measured.

The analysis of infectious diseases represents by far the largest clinical research area for e-nose technology attracting

TABLE I
DETECTION THRESHOLD LEVELS OF HUMAN OLFACTORY SYSTEMS AND ELECTRONIC NOSES

Volatile	Reported human threshold (ppm)	E-nose threshold (ppm)
Ethyl acetate	7-17	5-25
Butyric acid	0.4–10	<1
Diacetyl	$(4–15) \times 10^{-3}$	$(50–100) \times 10^{-3}$
n-Hexanal	$(10–50) \times 10^{-3}$	$(10–50) \times 10^{-3}$
Methional	$(2–50) \times 10^{-3}$	$(10–50) \times 10^{-3}$
Furanola	$(20–40) \times 10^{-6}$	$(50–100) \times 10^{-6}$
n-Nonane	0.2–7	<0.2
n-Octane	3–9	0.6
n-Heptane	7–13	<2
n-Hexane	13–30	<10
n-Pentane	20–50	40
1-Pentanol	0.13–1.3	<0.06
1-Butanol	0.2–1.3	0.3
1-Propanol	0.9–1.9	1.3
Ethanol	5–500	2
Methanol	13–600	3
Acetone	141	–
Ethanethiol	0.1×10^{-3}	–

–: not detected (when response <3 × back ground noise) at the same concentration as submitted to human noses.

both academic and commercial interest [5], [6]. However, given the low repeatability of the data patterns extracted from these sensors and the fuzzy nature of odor patterns, the use of advanced soft computing intelligence techniques is considered a necessity for accurate diagnosis.

The objectives of this study are to: 1) analyze 45 specimens of human urine *in vitro* by the application of an intelligent diagnostic model based on novel generation, detection, and rapid recognition of urinary volatile patterns within 5 h of receipt of specimens in the laboratory; 2) develop an extended normalized radial basis function (ENRBF) classifier trained using the expectation maximization (EM) algorithm and a novel application of the split and merge (SM) technique to neural networks (NNs) and evaluate its performance; 3) implement a multiple-classifier scheme, for nonlinear pattern recognition problems involving large and noisy data and explore the benefits of the fuzzy integral as a soft fusion method.

II. INSTRUMENT AND EXPERIMENTAL CONDITIONS

A. Urine Samples and Volatile Generating Kits

Approximately 80% of uncomplicated urinary tract infections (UTIs) are caused by *Escherichia coli* and 20% by enteric pathogens such as *Enterococci*, *Klebsiellae*, *Proteus* spp., coagulase (–) *Staphylococci*, and fungal opportunistic pathogens such as *Candida albicans*. Enhanced concentrations of fatty acids (especially acetic acid) and/or ammonia and amines have also been inspected. Positive to UTI samples (defined as being $>10^5$ cfu.ml^{-1}) ranged from 8 to 500 ppm acetic acid, with the average value being approximately 100 ppm. Using conventional methods, laboratory examination of urine requires 24 h incubation in order to obtain an accurate colony count.

TABLE II
MONOMERS AND DOPANTS USED IN THE BLOODHOUND BH114 SYSTEM

Sensor	Polymer monomer	Dopant
1	Ethylaniline	Sulphate (SO_4 counter ion)
2	Aniline	Sulphate (SO_4 counter ion)
3	Aniline	Sulphate (SO_4 counter ion)
4	Tryptophan	Sulphate (SO_4 counter ion)
5	Tryptophan	Sulphate (SO_4 counter ion)
6	2-methoxy-5- nitroaniline	Sulphate (SO_4 counter ion)
7	Thiophene and thiophene -3-tetrabutyl-ammonium carboxylic acid co-polymer	Perchlorate counter ion
8	Aniline Ethanol	Sulphate (SO_4 counter ion)
9	Pyrrole	Octanoic acid ethyl ester
10	Pyrrole	CO_4 counter ion
11	1, 4-phenlene diamine	Chloride counter ion
12	1, 4-phenlene diamine	Chloride counter ion
13	Pyrrole	Chloride counter ion
14	Pyrrole	Tetrabutyl-ammonium perchlorate

An additional 12 ± 24 h is needed for organism identification and susceptibility testing, which may further delay the administration of the most appropriate narrow-spectrum antibiotic. Forty-five 5 ml urine samples (following eukaryotic cell filtering extraction) were collected from randomly selected patients admitted in Gloucestershire Public Health Leadership Society (PHLS) and inoculated into specially made centrifuge bottles (50 ml) each containing 95% brain heart infusion (BHI) broth, 5% serum bovine, 0.70 mg.ml^{-1} of a series of amino acids, 1 mg.ml^{-1} urea, 0.75 mg.ml^{-1} lactose, 0.1 mg.ml^{-1} casein, 0.3 mg.ml^{-1} acetylcholine to a final volume of 20 ml per volatile generating kit (VGK) and incubated aerobically for 5 h at 37 °C [7].

B. Gas Sensing System and Flow Injection Analysis (FIA) of Urinary Volatiles

The e-nose (Bloodhound BH-114) used in this study employed 14 conducting polymer sensors. Specific selection and polymer tailoring, doping materials, and precise manufacturing process can make each of the 14 sensors consistently responsive to a variety of volatile mixtures. Table II illustrates the monomers and dopants used for BH114 device. The BH114 instrument consists of a fully integrated sampling system with the sensor array and hardware controlled by proprietary Windows-based software, which also incorporates data collection and preprocessing software (Fig. 1) [8]. The interaction of the VOCs and the conducting polymer surface produces a change in resistance, which is measured and subsequently displayed on the computer screen. After 5 h of incubation to coincide with the logarithmic phase of growth, 45 VGK were placed in a 37 °C water bath and directly connected to the e-nose by inserting a needle into the headspace of the sample vials. The actual urine sampling time and baseline recovery per specimen was 3 min. The individual samples were analyzed in a random order. A data capture software analyzed complex volatile patterns liberated over the headspace of 45 UTI-VGK specimens and extracted multiple sensor parameters from each patient profile for further postprocessing analysis.

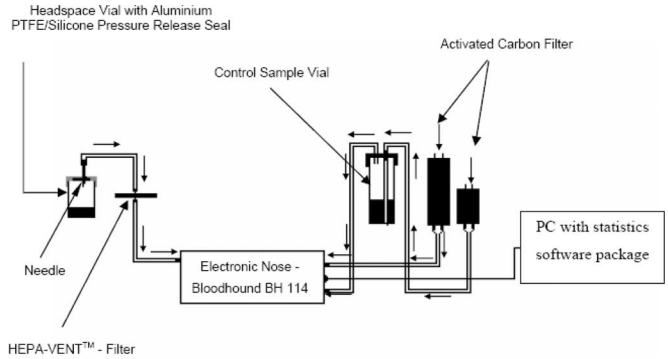

Fig. 1. Overview of e-nose configuration.

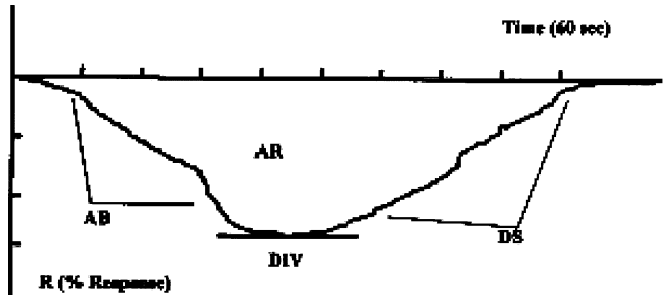

Fig. 2. Parameters measured for each sensor response.

III. Data Exploration

From the 45 randomly selected samples, 30 cases of UTI were identified using standard microscopy and culture. Of these, 13 were infected with *E. coli* (e), nine with *Proteus* sp. (p), and eight with coagulase–*Staphylococcus* sp. (st). The remaining ones were clean cases (n). These data were randomly divided into two groups. The first composed of 31 samples (e: 9, p: 6, st: 5, n: 11) that were used as a "training" group. The second set of 14 samples (e: 4, p: 3, st: 3, n: 4) was used as a validation set. The sensor responses from each of the samples were analyzed through the extraction of four sensor features that the sensor–volatile physiochemical interaction and pattern extraction. These features are: 1) divergence (DIV): maximum step response; 2) absorption (AB): maximum rate of change of resistance; 3) desorption (DS): maximum negative rate of change of resistance; 4) area under the curve (AR). Fig. 2 displays a real-time sensory response curve taken from a microbial volatile headspace [9]. The 14 conducting polymers and the four sensor features thus provided a set of 56 sensor parameters extracted for each sample.

e-Nose raw data might not be the most useful input variables for statistical analysis due to noise, sensor drift, or inconsistency. In our case, responses have been normalized according to

$$x'_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2}} \tag{1}$$

where x'_{ij} is the normalized sensor response, x_{ij} is the raw sensor response, and x_{ik} are the individual sensor responses.

A. Multiple Neural Network Fusion

In this study, four subsystems have been developed, and each of them was associated with the four parameters extracted from the sensor response curve. Each feature has been modeled with a 14-input 4-output subnetwork. This provides a degree of certainty for each classification based on the features' own inputs. The outputs of each of these networks must then be combined to produce a total output for the system as a whole. The aim in this study is to incorporate information from each feature space so that decisions are based on the whole input space. The simplest method, the average, does not take into account the objective evidence supplied by each of the feature classifiers and confidence that we have in that classifiers results.

The fuzzy integral is an alternative method that claims to resolve both of these issues by combining evidence of a classification with the systems expectation of the importance of that evidence. By treating the classification results, a series of disjointed subsets of the input space Sugeno defined the g_λ-fuzzy measure [10] as

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)\, g(B), \quad \lambda \in (-1, \infty) \tag{2}$$

where the λ measure can be given by solving the following nonlinear equation:

$$\lambda + 1 = \prod_{i=1}^{K} \left(1 + \lambda g^i\right), \qquad \lambda > -1. \tag{3}$$

When combining multiple NNs, let g^i denote the fuzzy measure of network *i*. These measures can be interpreted as quantifying how well a network properly classified the samples/patterns. They must be known and can be determined in different ways, i.e., the fuzzy measure of a network could equal the ratio of correctly classified patterns during supervised training over the total number of patterns being classified. In this research, each network's fuzzy measure equaled $1 - K_i$, where K_i was network *i*th overall testing kappa value [11]. A pattern is being classified to one of *m* possible output classes, c_j for $j = 1, \ldots, m$. The outputs of *n* different networks are being combined, where NN_i denotes the *i*th network. First, these networks must be renumbered/rearranged such that their *a posteriori* class probabilities are in descending order of magnitude for each output class *j*

$$y_1(c_j) \geq y_2(c_j) \geq \cdots \geq y_n(c_j)$$

where $y_i(c_j)$ is the *i*th network's *a posteriori* class *j* probability. Next, each network/set of networks' g_λ-fuzzy measure is computed for every output class *j* and is denoted by $g_j(A_i)$. $A_i = \{\mathrm{NN}_1, \mathrm{NN}_2, \ldots, \mathrm{NN}_i\}$ is the set of the first *i* networks ordered correspondingly to class *j's* associated *a posteriori* probabilities. These values can then be computed using the following recursive method.

$$\begin{aligned} g_j(A_1) &= g_j(\{\mathrm{NN}_1\}) = g^1 \\ g_j(A_i) &= g_j(\{\mathrm{NN}_1, \ldots, \mathrm{NN}_i\}) \\ &= g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}), \quad \text{for } 1 < i < n \\ g_j(A_n) &= g_j(\{\mathrm{NN}_1, \ldots, \mathrm{NN}_n\}) = 1. \end{aligned}$$

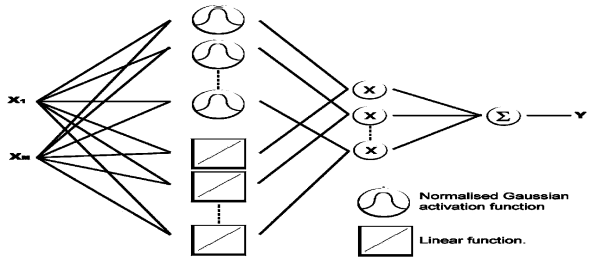

Fig. 3. ENRBF concept.

Finally, the fuzzy integral for each class *j* is defined as [12]

$$\max_{i=1}^{n}[\min[y_i(c_j),\ g_j(A_i)]]. \tag{4}$$

The class with the largest fuzzy integral value is then chosen as the output class to which the pattern is classified.

IV. Learning-Based Classifiers

Each subsystem in the proposed multiple-classifier concept was modeled with an appropriate intelligent learning scheme. Locally active networks are less prone to dimensionality problems as each single neuron node defines a multidimensional hypersphere in the input domain. This hypersphere is often defined through the use of a Gaussian activation function placed on the neuron. Each neuron has a mean parameter and a deviation parameter. The mean parameter defines the location of the neurone in the input domain while the deviation parameter defines the size and the shape of the neuron. If a single value is used as the deviation parameter, then the neuron has an equal size along each axis of the input domain [13]. More frequently, a diagonal or full variance matrix is instead used. This alternative way provides differing levels of deviation along each axis of the input domain. Networks utilizing these types of neurones are often called normalized radial basis function (NRBF) networks. These are computationally similar to Gaussian mixture models (GMMs) and methods for fixing the parameters of one can often be applied to the other.

The proposed modified extended NRBF (ENRBF) extends the NRBF network by incorporating a linear sum associated with each neuron [14]. In this study, we consider the neurons of the ENRBF as a GMM so as to be able to readily apply the EM algorithm. The proposed scheme is illustrated in Fig. 3. For a system with *N* sets of *C* inputs contained in a matrix *X*, *L* outputs, and *M* partitions/neurones, we can define the output $y_1, \ldots, L$ for an input vector *x* as

$$\hat{y}_l\,(x_i) = \sum_{j=1}^{M} \Phi_j\,(x_i)\,\mathrm{A}_j^l\,(x_i). \tag{5}$$

For each output $l = 1, \ldots, L, l = 1, \ldots, N$,

$$\Phi_j(x_i) = \frac{\Psi_j(x_i)}{\sum_{j=1}^{M} \Psi_j(x_i)} \tag{6}$$

$$\Psi_j(x_i) = \alpha_j \frac{1}{(2\pi)^{M/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu_j)^T(x_i-\mu_j)}{\Sigma_j}\right) \tag{7}$$

where $\Theta_j = (\alpha_j, \theta_j),\ \theta_j = (\Sigma_j, \mu_j)$ are parameters of the network, and

$$A_j^l(x_i) = a_j^{lT} x_i \tag{8}$$

where a^l and $x_{i,0} = 1$. There are two sets of parameters to be modified through learning, i.e., Θ parameters that govern the position and size of the neuron. Using the aforementioned notation, α is a scaling factor or prior probability in a GMM, μ is the center or mean value, and Σ is the diagonal or complete covariance matrix representing the deviation along each axis of the input domain. The parameters a in (8) are the coefficients of the linear model.

A. EM Algorithm for GMM and Linear Models

The EM algorithm is an iterative approach that aims to compute a new parameter set Θ based on the maximization of a likelihood function describing Θ with respect to a data set *X*. The purpose of applying an unsupervised clustering process such as the EM algorithm to GMM is to determine the mean, deviation, and prior probability of the GMM that represents the locally active neurons of the ENRBF. The distribution of the neurons can be represented by the following probabilistic model:

$$p(x|\Theta) = \sum_{i=1}^{M} \Psi_i(x) = \sum_{i=1}^{M} \alpha_i p_i(x|\theta_i). \tag{9}$$

The equations for updating the parameters at each step are calculated as [15]

$$\alpha_i^{(t)} = \frac{1}{N}\sum_{n=1}^{N} \Phi_i(x_n) \tag{10}$$

$$\mu_i^{(t)} = \frac{\sum_{n-1}^{N} x_n \Phi_i(x_n)}{\sum_{n-1}^{N} \Phi_i(x_n)} \tag{11}$$

$$\Sigma_i^{(t)} = \frac{\sum_{n=1}^{N} \Phi_i(x_n)(x_n - \mu_i^{(t)})(x_n - \mu_i^{(t)})^T}{\sum_{n=1}^{N} \Phi_i(x_n)}. \tag{12}$$

Starting from an initial set of values, these equations are guaranteed to converge to a local maximum value for the $Q(\Theta^{(t)}|\Theta^{(t-1)}) = \sum_{n=1}^{N}\sum_{n=1}^{M}\{\log \alpha_i p_i(x_n|\theta_i)\} \times P(i|x_n, \Theta^{(t-1)})$ function [12]. The GMM provides a means to define local areas within the input space. Each density in the mixture represents a neuron in the ENRBF structure, the normalized probability of a given *x* belonging to given *i*, a specific density is the neuron activation level Φ_i for $i \in \{1, \ldots, M\}$. The linear coefficients of the local models can be identified by adopting the assumption that the parameters of the neurons are fixed. This is carried out by explicitly declaring a variable representing the expectation of the complete data as a function of the incomplete data *X*. Following the two-step algorithm, the expectation step provides a desired output from each neuron–local model pair. The maximization step then optimizes the linear coefficients of $A(\cdot)$ so as to fit each of the linear models to the members of *X* covered by each neuron. The parameters for the next iteration are then calculated by

$$a_i^{l(t)} = (X^T\,\Phi_i X)^{-1}(X^T\,\Phi_i u_i^l) \tag{13}$$

for $i \in \{1, \ldots, M\}$ and $l = 1, \ldots, L$, where

$$u_i^l = [\, u_{i,1}^l \quad \cdots \quad u_{i,N}^l \,]^T$$
$$\Phi_i = \text{diag}\,(\Phi_i(x_{i,1}), \ldots, \Phi_i(x_{i,N})).$$

The weighting factor Φ_i, $i \in \{1, \ldots, M\}$, used in order to insure that local models are weighted toward those input patterns that are within the neurons' activation area [16].

B. Split and Merge EM (SMEM) Algorithm

The EM algorithm is guaranteed to train the parameters to a local optimum. However, the EM process for the identification of the parameters for the Gaussian activation functions is very dependant on the choice of initial values. Without very careful selection of initial values, it is highly likely that, after training, there will be some areas of the input domain that are overly populated with activation neurons and some areas not sufficiently represented. A set of criteria has been proposed that can be used to identify neurons in overpopulated domains to be merged, and neurons that do not adequately describe the area of the input domain they cover to be split [17]. Standard EM is performed until convergence is satisfied. Two neurons are selected for merging using the following matching criteria:

$$J_{\text{merge}}(i, j|\Theta^{(t)}) = \Phi_i^T \Phi_j \tag{14}$$

where

$$\Phi_i = \lfloor \Phi_i(x_{i,1}), \ldots, \Phi_i(x_{i,N}) \rfloor,$$
$$i \in \{1, \ldots, M\}, \qquad i \in \{1, \ldots, M\}, \qquad i \neq j.$$

Two components with a high value can be considered as good candidates for merging. The split criteria are performed through the use of the local Kullback divergence that gives a measure a degree of entropy between two distributions. A measure is taken between the local divergence around a specific neuron $g_k(\cdot)$ and the empirical distribution across the whole data space $h_k(\cdot)$:

$$J_{\text{split}}(k|\Theta^{(t)}) = \sum_{n=1}^{N} g_k(x_n|\Theta^{(t)}) \ln \left[\frac{g_k\left(x_n|\Theta^{(t)}\right)}{h_k\left(x_n|\Theta^{(t)}\right)} \right]$$
$$g_k(x|\Theta^{(t)}) = \frac{\sum_{n=1}^{N} \delta\,(x - x_n)\,\Phi_k\,(x_n)}{\sum_{n=1}^{N} \Phi_k\,(x_n)}$$
$$h_k(x|\Theta^{(t)}) = \frac{1}{N}\delta\,(x - x_n) \tag{15}$$

for $k \in \{1, \ldots, M\}$ and where $\delta\,(\cdot)$ is the Kronecker delta function. The *k* with the largest value is a candidate for splitting. To perform merging operations, the selected neurons *i* and *j* are combined by the following process:

$$\alpha_i^{(t+1)} = \alpha_i + \alpha_j \tag{16}$$

$$\theta_i^{(t+1)} = \frac{\theta_i^{(t)} \sum_{n=1}^{N} \Phi_i(x_n) + \theta_j^{(t)} \sum_{n=1}^{N} \Phi_j(x_n)}{\sum_{n=1}^{N} \Phi_i(x_n) + \sum_{n=1}^{N} \Phi_j(x_n)}. \tag{17}$$

This leaves the neuron previously indexed by *j* free to be used for one half of the split neuron. As a result, the number of neurons in the system remains constant. The "split" operation in this study is performed by splitting the neuron along its largest axis using the eigenvalues of the *k*th mixture [18]. By putting together the SMEM method for GMM and the EM-based technique for linear parameter estimation, the following algorithm has been implemented. The goal of this is to establish both the locations and deviations of the Gaussian components and the linear models [18].

Step 1: Initialize parameters for network.
Step 2: Perform standard GMM EM operations until convergence using (10)–(12). Identify parameters for local models using (13).
Step 3: Calculate merge candidates using (14).
Step 4: Select best candidates for merge setting the value in J_{merge} to 0.
Step 5: Calculate split candidates using (15).
Step 6: Select best candidate for split setting the value in J_{split} to 0.
Step 7: Perform SM, then the partial EM steps. Then select parameters for the linear models using (13), and call this parameter set Θ^*.
Step 8: If results of $\Theta^{(t)}$ are better than results of Θ^*, then $\Theta^{(t)} = \Theta^*$, and go back to step 2). Else if results of $\Theta^{(t)}$ are worse or the same as results of Θ^* and there are further values of J_{split}, go back to step 6). Else if results of $\Theta^{(t)}$ are worse or the same as results of Θ^* and there are further values of J_{merge}, go back to step 4).
Step 9: Finish with the best parameter set $\Theta^{(t)}$.

V. Discussion of Results

The adopted architecture reduced the dimensionality of the network search space from 56 to 14 because the input space of each module is divided over the four nets ($4 \times 14 = 56$). Reducing the search space in this way increases both computational efficiency and the probability that optimal network parameters will be found within the search space. The accuracy of a high-dimensional network is also directly related to an increased number of training samples. For the case of complex clinical patterns, it is inevitable to apply NNs due to the high variability between the individual samples [19]. Each class must be composed of representative and reproducible samples. In the case of clinical cases, it is usually difficult to obtain a massive number of samples from the National Health Service (NHS) laboratories, and emphasis is given to the quality of data rather than the quantity [5], [8]. Each of the ENRBF subnetworks consisted of four outputs corresponding to each of the classes and 14 inputs, while six local neurons have been used with the corresponding linear models. The parameters for the activation functions were initialized by selecting random values in the range of the entire distribution. The parameters for the linear models were initialized with random values between 0 and 1. Each network was trained in their respective feature space. Although the decision boundary for a two-class problem is where the posterior probabilities of these two classes equal 0.5, in the case of multiclass problems, there is an increase in uncertain areas, as these boundaries can be in regions with posterior probabilities between 1/*C*

TABLE III
FEATURES ACCURACY FROM INDIVIDUAL SUBNETWORKS

Modules	Accuracy (14 testing patterns)
Divergence	92.86%(1 mistake)
Absorption	64.28% (5 mistakes)
Desorption	85.71% (2 mistakes)
Area	71.43% (4 mistakes)
Overall	100% (0 mistakes)

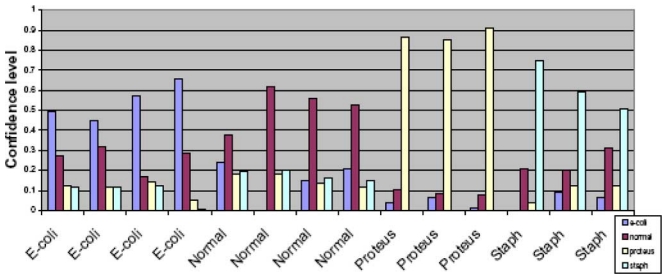

Fig. 4. Performance using average as a fusion method.

and 0.5, where C is the number of classes. This also increases the chance that the classification will change when a fusion combination rule will be used. It is also important that all class probabilities should be calculated for specific data [20]. When a higher confidence is required for classification tasks, especially in clinical cases, an acceptance threshold on the class probability can be introduced. Then probabilities around the decision boundary are excluded from classification. In that case, the decision obtained by the multiple-classifier scheme is judged by the medical consultant who will take the final decision.

Divergence achieved correct classification in 13 of the 14 "testing" patterns incorrectly classifying one of *E. coli* infected samples as normal. The minimum decision boundary was 0.53 showing that the quality of the classification was high. The subnetwork trained to classify the absorption data correctly classified 9 out of 14. The minimum boundary was however 0.349. This is clearly above the 0.25 threshold defined for a four-class problem. The network trained on Desorption data incorrectly classified two *E. coli* samples as normal. However, the minimum decision boundary was 0.54 that shows that those classified were done confidently. The network trained to classify the area readings classified 10 out of the 14 "testing" examples correctly. It misclassified two *Staphylococcus* as normal and two normal cases as being infected with *E. coli.* The related decision level for the correctly classified cases was 0.4277. These results are shown in Table III.

These results were fused using the two fusion methods described in the previous section. By averaging the outputs from each of the four feature-based networks, an overall accuracy of 100% was achieved. As can be seen in Fig. 4, decision levels of the diagnosis drop as low as 0.3776 for the case of one normal sample. The decision for *E. coli* for this specific sample was 0.24. This is clearly the case where medical support is needed for the final decision.

These results are also verified through the calculation of the parameters normally used in medical diagnostics tests [21]: sensitivity, 88%; specificity, 60%; and predictability, 80%. How-

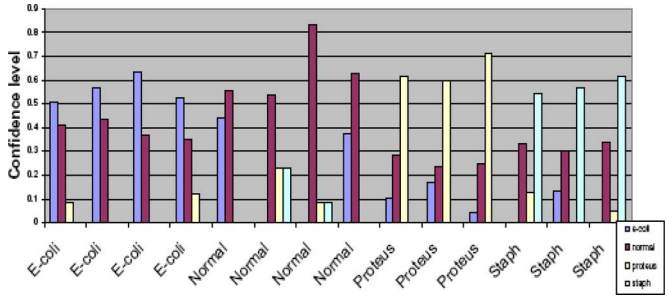

Fig. 5. Performance using fuzzy integral as a fusion method.

ever, by applying the fuzzy integral methodology as a fusion scheme, as can be seen in Fig. 5, the accuracy for the system as a whole is still 100%, but now all of the decision levels for the diagnosis are over 0.5. These results are also verified through the calculation of the remaining parameters: sensitivity, 100%; specificity, 100%; and predictability, 100%.

RBF networks have a clear advantage over multilayer perceptrons, in terms of accuracy and faster training times. However, the problem of determining RBF's structure in an optimal way still exists. The ENRBF can be considered as an extension of the classic RBF simple architecture. The approach taken here attempts to address some of the established problems with EM training of Gaussian systems. The SMEM algorithm has been incorporated into the ENRBF network. The inclusion of the SM concept attempts to tackle the problem of identification of local maxima when calculating the parameters of the Gaussian functions. The novel application of the SMEM approach has shown to provide an effective means of overcoming the problem of local optimum within EM training for the selection of neurons parameters. The use of eigenvalues and eigenvectors to calculate the axis of a neuron split operation instead of random values has been adopted. The ENRBF model has been extended with the least squares EM algorithm for fixing the parameters of the linear models. We have also made the use of the information contained in the covariance matrix when carrying out the split operations. This is to improve the stability of the EM algorithm and to also mean that the experiments are repeatable. The application demonstrates that the technique, which is characterized by its speed in training processing, is capable of modeling complex high-dimension data.

VI. CONCLUSION

There is an urgent need for new point-of-care devices that are not only cheap and robust but also simple to operate. In this study, an alternative approach based on gas-sensing technology was taken to investigate the suitability of such a system as a point-of-care device. It should be emphasized here that this system is not being proposed as a replacement for a clinician's diagnosis but rather to supplement other diagnostic methods. It also helps the clinician deliver better service as the e-nose system has the potential advantage of making decisions 24 h per day, seven days per week. This study suggests that the e-nose combined with advanced learning-based processing tools is able to identify specific bacterial pathogens with accuracy and speed,

even with a small sample quantity, at the point-of-care. Chronic renal failure and tuberculosis are also two diseases where people could benefit from new point-of-care devices based on gas sensors.

REFERENCES

[1] V. Moret-Bonillo, "Integration of data information and knowledge in intelligent patient monitoring," *Expert Syst. Appl.*, vol. 15, pp. 155–163, 1998.

[2] M. Phillips, "Method for the collection and assay of volatile organic compounds in breath," *Anal. Biochem.*, vol. 247, pp. 272–278, 1997.

[3] S. Ampuero and J. Bosset, "The electronic nose applied to dairy products: A review," *Sens. Actuators B*, vol. 94, pp. 1–12, 2003.

[4] W. Gopel, "Chemical imaging: I. Concepts and visions for electronic and bioelectronic noses," *Sens. Actuators B*, vol. 52, pp. 125–142, 1998.

[5] K. Persaud, A. M. Pisanelli, P. Evans, and P. Travers, "Monitoring urinary tract infections and bacterial vaginosis," *Sens. Actuators B*, vol. 116, pp. 116–120, 2006.

[6] A. K. Pavlou, V. S. Kodogiannis, and A. P. F. Turner, "Intelligent classification of bacterial clinical isolates in vitro, using electronic noses," in *Proc. Int. Conf. Neural Netw. Expert Syst. Med. HealthCare*, 2001, pp. 231–237.

[7] A. K. Pavlou, N. Magan, D. Sharp, J. Brown, H. Barr, and A. P. F. Turner, "An intelligent rapid odour recognition model in discrimination of *Helicobacter pylori* and other gastroesophageal isolates in vitro," *Biosens. Bioelectron.*, vol. 15, pp. 333–342, 2000.

[8] R. Fend, C. Bessant, A. J. Williams, and A. C. Woodman, "Monitoring haemodialysis using electronic nose and chemometrics," *Biosens. Bioelectron.*, vol. 19, no. 12, pp. 1581–1590, 2004.

[9] V. S. Kodogiannis, A. K. Pavlou, P. Chountas, and A.P. F. Turner, "Evolutionary computing techniques for diagnosis of urinary tract infections *in vivo*, using gas sensors," in *Neural Computing and Soft Computing* (Advance in Soft Computing), New York, 2003, pp. 474–479.

[10] M. Sugeno, "Fuzzy measures and fuzzy integrals: A survey," in *Fuzzy Automata and Decision Processes*, M. M. Gupta, G. N. Saridis, and B. R. Gaines, Eds. Amsterdam, The Netherlands: North Holland, 1977, pp. 89–102.

[11] S. Mitra, S. K. Pal, and P. Mitra, "Data mining in soft computing framework: A survey," *IEEE Trans. Neural Netw.*, vol. 13, no. 1, pp. 3–14, Jan. 2002.

[12] L. I. Kuncheva, *Fuzzy Classifier Design*. Heidelberg, Germany: Physica-Verlag, 2000.

[13] S. Chen, A. Billings, and P. M. Grant, "Orthogonal least squares algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.

[14] E. Wadge, V. S. Kodogiannis, and D. Tomtsis, "Neuro-fuzzy ellipsoid basis function multiple classifier for diagnosis of urinary tract infections," in *Proc. ICCMSE 2003*. Kastoria, Greece, pp. 673–677.

[15] E. Alpaydin, "Soft vector quantization and the EM algorithm," *Neural Netw.*, vol. 11, no. 3, pp. 467–477, 1998.

[16] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[17] N. Ueda and R. Nakano, "EM algorithm with split and merge operations for mixture models," *IEIC Trans. Inf. Syst.*, vol. 83, no. 12, pp. 2047–2055, 2000.

[18] E. Wadge, "The use of EM-Based neural network schemes for modelling and classification," Ph.D. dissertation, Westminster Univ., London, U.K., 2005.

[19] R. Fend, "Development of medical point-of-care applications for renal medicine and tuberculosis based on electronic nose technology," Ph.D. dissertation, Cranfield Univ., Cranfield, U.K. 2004.

[20] D. Tax, M. van Breukelen, R. Duin, and J. Kittler, "Combining multiple classifiers by averaging or by multiplying?," *Pattern Recognit.*, vol. 33, pp. 1475–1485, 2000.

[21] P. Szczepaniak, P. Lisboa, and J. Kacprzyk, *Fuzzy Systems in Medicine*. New York: Springer-Verlag, 2000.

Vassilis S. Kodogiannis (M'01) received the Elec. Eng. degree from Democritus University of Thrace, Xanthi, Greece, in 1990, the M.Sc. degree in very large-scale integration (VLSI) systems engineering from the University of Manchester Institute of Science and Technology (UMIST), Manchester, U.K., in 1991, and the Ph.D. degree in electrical engineering from Liverpool University, Liverpool, U.K., in 1994.

He is currently a Principal Lecturer in the School of Computer Science, University of Westminster, London, U.K. His current research interests include the areas of intelligent systems, image processing, and control.

Dr. Kodogiannis is a member of the Technical Chamber of Greece.

John N. Lygouras received the Elec. Eng. degree and the Ph.D. degree in electrical engineering from Democritus University of Thrace, Xanthi, Greece, in 1981 and 1990, respectively.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, Democritus University of Thrace. His current research interests include the areas of robotics, image processing, and analogue/digital electronic circuits.

Andrzej Tarczynski (M'01) received the M.Eng. and Ph.D. degrees in automatic control from Warsaw University of Technology, Warsaw, Poland, in 1979 and 1986, respectively.

He is currently a Reader in signal processing and control systems in the Department of Electronic Systems, University of Westminster, London, U.K. His current research interests include the areas of signal processing and control systems.

Hardial S. Chowdrey received the Ph.D. degree from St. Thomas' Hospital, Medical School London, London, U.K., in 1984.

He is currently a Professor and the Head of the Department of Biomedical Sciences, School of Biosciences, University of Westminster, London. His current research interests include the areas of pharmacology, physiology, and biomedical sciences.