Identification of bacterial pathogens using quadrupole mass spectrometer data and radial basis function neural networks

J.W.T. Yates, J.W. Gardner, M.J. Chappell and C.S. Dow

Abstract: A quadrupole mass spectrometer has been employed to analyse the headspace above bacterial cultures. This, along with a pattern recognition algorithm, constitutes an electronic nose system. Here we present the results of a study on the headspace of pathogens, specifically *Escherichia coli K12* and *Staphylococcus aureus*, the purpose being to identify the growth phase and strain of different pathogens. The data collected from the mass spectrometry were used to train a radial basis function (RBF) neural network. This type of network was employed because it requires smaller training sets and is suitable for what is, in effect, 505 mass 'sensors'. Principal components analysis shows that there is sufficient information in the volatiles to discriminate between the different growth phases of *E. coli*, but less so for two strains of *S. aureus*, i.e. MRSA and NCTC. Excellent results are obtained using these RBF neural networks as approximaters of discriminant functions. Furthermore, it is demonstrated that this method can deal with classification problems that involve nonlinearity in the data. It is concluded that the reported methodology shows promise as a useful pathogen identification technique, and in particular discrimination between the virulent MRSA and the innocuous NCTC strain.

1 Introduction

Due to the emergence of antibiotic resistant micro-organisms, the rapid screening of microbial pathogens has become a crucial issue. *Staphyloccocus aureus* is a serious human pathogen responsible for many cases of septicaemia and toxic shock syndrome [1]. Methicillin resistant *Staphylococcus aureus* (referred to as MRSA) was first reported in 1961, soon after the antibiotic methicillin entered clinical use, and the pathogen is now becoming a major problem in hospitals. This pathogen is also responsible for mastitis in cows and sheep, with severe economic repercussions.

Not only has it become apparent that identification of the strain of the organism is important, but it is also necessary to be able to identify its metabolic state. This relates to the viability of a microbe and hence its response when challenged with antibiotics. Moreover, micro-organisms, including certain strains of *Escherichia coli*, display temporal expression of particular genes [2]. In the case of *E. coli O157*, responsible for many cases of food poisoning, verocytotoxin gene expression is enhanced in dormant cells, i.e. at a low growth rate stage.

IEE Proceedings online no. 20041145

doi:10.1049/ip-smt:20041145

Paper first received 19th February 2003 and in revised form 9th September 2004. Originally published online: 11th April 2005

J.W.T. Yates is with AstraZeneca R&D, 8AF2, Mereside, Alderley Park SK10 4TG, UK

J.W. Gardner and M.J. Chappell are with the School of Engineering, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

C.S. Dow is with the Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

J.W.T. Yates was formerly with the School of Engineering, University of Warwick

Email: James.Yates@astrazeneca.com

We use the term headspace analysis to encompass the various technologies used for the detection of volatile compounds and the resulting data processing.

Headspace analysis holds a great deal of potential for medical applications, see for example [2, 3] and [9]. The emphasis to data has been on its ability to perform swifter diagnosis. The analysis of an odour takes only a few minutes (approximately 5 min) in comparison to hospital pathology labs, which can take several days to culture and identify a pathogen. Therefore, chemical headspace analysis could represent a real advancement in terms of time, efficiency and cost.

We concern ourselves here with volatile analysis using a quadrupole mass spectrometer, specifically an Agilent 4440. This instrument determines the abundance of molecules in a sample over the mass range of 46 to 550 daltons. We investigate whether it is possible to distinguish between different samples by detecting only the volatiles that the organisms give off.

Neural networks have proved useful methods both for pattern matching and time series analysis. The difficulty with approaches such as single and multilayer perceptron methods is that increases in complexity of the network must be matched with an increase in the size of the training data set. This becomes particularly apparent for high-dimensional data, for example mass spectrometry data. In this situation the dimensionality of the input space is dictated by the mass range of the equipment, i.e. a difference of 505 daltons.

Thus, an immediate problem with this method of headspace analysis, from the perspective of neural networks, is the demand for large training sets. This curse of dimensionality dictates that, for Agilent 4440 data, training samples of the order of a thousand points are required to produce a sufficiently 'dense' set of examples [5, 6]. Radial basis function (RBF) networks reduce this problem via

a priori assumptions relating to smoothness of the discriminant function [7, 8].

Here we consider two different types of data set. We do this to demonstrate that the techniques outlined in this work can detect both structural and metabolic changes in a cell without knowledge of the chemical constituents of a bacterial sample; these would be difficult to identify using a mass spectrometer given the complex mixture used. The first data type consists of samples of two strains of *Staphyloccocus aureus*. One is antibiotic resistant, the other is not and we refer to these as MRSA and NCTC, respectively. The second data type consists of *E. coli* cultures measured at different growth stages. This is also used to show how binary classifiers can be easily extended to n-classifiers.

2 Data collection

The data were collected in the Biological Sciences Department at the University of Warwick and were used for a previous study [2]. Headspace samples were formed by the injection of pure helium gas into a 25 ml vial containing 10 ml of culture, followed by robotic transfer of the sample vial into a heated stage and its stabilisation at a temperature of 37($\pm$0.1)°C. The headspace was then injected into a quadrapole mass spectrometer. The headspace autosampler had a repeatability of about 0.25% by volume. The mass spectrometer analysed the mass content of the headspace with the range set to 46 to 550 daltons and a resolution of 0.1 daltons. The spectrometer could record individual masses but the abundance was typically thousands of mass units. The unit also contained a series of internal diagnostics to check ion gauge currents, vacuum levels and so on. Overall, we estimate the repeatability of the system to be within 1% over the period of operation. Figure 1 illustrates the experimental set up.

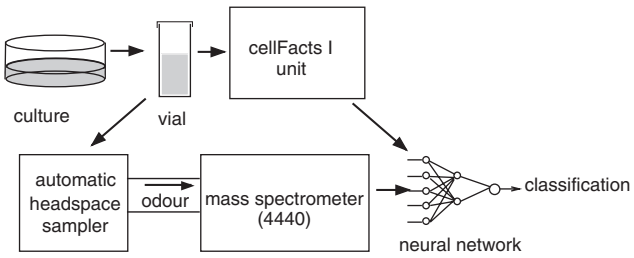

Fig. 1 *Schematic of sample and data flow*

The samples were classified in the laboratory using established methods. For the *S. aureus* data this was a case of checking whether a sample consisted of the required strain. This was easily checked as separate cultures of the two strains were grown. The samples were taken at hourly intervals, so the data also contained information about the growth state of the culture. Two sets of experiments were carried out yielding two data sets.

Microbial cells pass through three stages of growth. These may be subdivided, but here we consider the three main stages as these can be reliably observed (Fig. 2). In the 'lag' phase, the culture adjusts to its new environment. The population size, i.e. number of cells, remains stable but there is a dramatic change in the size of the individual cells comprising the population. In the 'log' phase, the culture undergoes growth and division. There is a shift in the cell size distribution which correlates with the growth rate of the individual cells. The third phase is the stationary phase, where most of the nutrient resources available to the culture have been used up, and so the cells enter into a form of stasis.

For the *E. coli* data, growth phases were identified via cell count and size [2], the actual data are displayed in Fig. 2. These were measured using a CellFacts I instrument, which measures the size distribution of particles in a liquid sample. For further background information on this system see [9]. Samples were drawn from the culture at hourly intervals for the first 8 h of the experiment and then at 24, 48 and 72 h.

The samples were introduced, at each time point, to the mass spectrometer. The Agilent 4440 heats the sample vial up to 80° for 3 min, and then draws off the resulting saturated static headspace. This vapour is then passed to a quadrupole mass spectrometer for analysis. The whole process takes approximately 5 min. For more information see [2].

The data output is in the form of a mass distribution (called abundance) in the range here of 46 to 550 daltons. The abundance of each mass is the count of particles of that mass in the sample. The output data were logged on a PC running software provided by the manufacturers of the Agilent 4440 and analysed to see whether it is possible to discriminate between the different classes of samples. Each data point is labelled using the classification achieved following the methods detailed below.

3 Principal components plots

Principal components analysis (PCA) is a data visualisation technique that searches for variation in the data provided. It

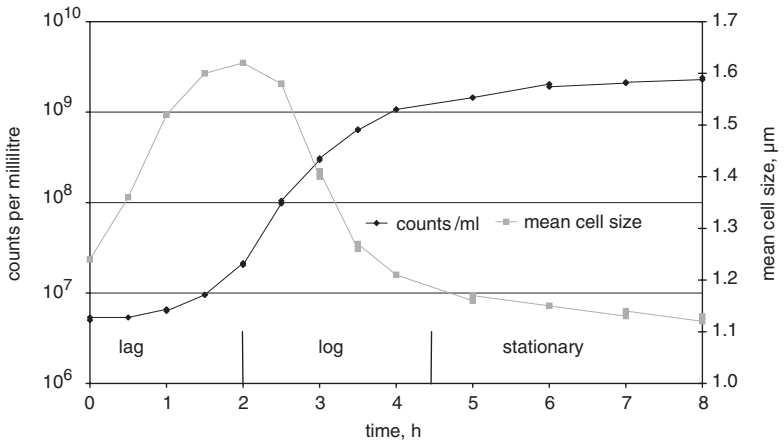

Fig. 2 *CellFacts data showing the change in cell and population size*

looks for 'directions' in the feature space along which the data predominantly vary. These principal components are ranked in order and in this study three are selected to provide plots for visual inspection. The criteria for selection in this study are greatest variance or greatest separation between classes. For further information see [10].

The PCA plots provided demonstrate the result of projecting the data onto the selected principal components. Thus, in 3-D, we obtain a representation of how the data are varying within and between classes.

PCA of the data sets collected was performed using Matlab version 6 software. Figure 3 shows a PCA plot obtained from one of the *S. aureus* data sets, here 'x' represents a MRSA sample and 'o' represents an NCTC sample. It can be seen that the classes are not linearly separable. We illustrate this with a plot where the principal components have been selected using cluster separating criteria derived from the statistical, or Mahalanobis, distance [11], which is a metric based upon the variance within the data set. We also observe two obvious outliers. These were not removed from the data set, on the grounds that the data sets were small.

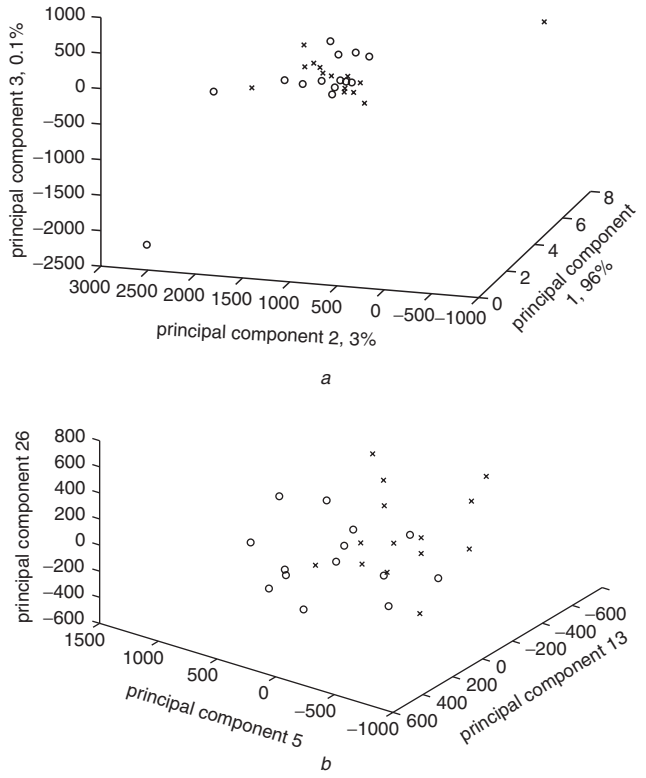

Fig. 3 *Principal component plots for S. aureus data where the crosses represent antibiotic resistant cultures*
a First three principal components (percentage variation is also shown). The data used were the raw output of the Agilent 4440. Notice that principal component 3 only describes 0.1% of the variance of the data. Hence, this plot illustrates the majority of the variance observed in this work.
b Selected principal components

Attempting to separate data clusters with a plot of the first three principal components is made on the implicit hypothesis that the only variance present is due to the differences between data classes [12]. In reality there are a number of reasons for variation in the data, for example the measuring equipment may have some built-in error. In addition samples that are used to estimate the parent distribution will not be perfect as, for example in the case of binary classification, samples will have alternative features that are not of interest to this study. For instance, when looking for structural differences in bacteria, there may be variation due to metabolic processes. These factors may result in the within-class variation being greater than that between classes.

It was therefore necessary to find an algorithm to detect which principal components describe the greatest separation between two classes. In [12], it is shown that the Mahalanobis distance between two classes, based upon the kth principal component, is a monotonic increasing function of the expression given in (1):

$$[V_k^{\mathrm{T}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^2/\lambda_k \tag{1}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the within class average vectors, V_k is the kth principal component and λ_k is the variance described in the direction of V_k.

This suggests a very useful criterion for selecting the 'best' principal components to plot. We simply seek to maximise (1). A three-component plot can be achieved by choosing the components which yield the highest values of (1) (as in Fig. 3*b*). In this way the distance is maximised between classes observed in the plot.

Looking again at the 3-D principal components plots, it is evident that the classes cannot be separated using a single hyperplane.

An examination of a PCA plot for the *E. coli* data in Fig. 4 shows that these data exhibit stronger clustering. However, there is still some overlap of classes. The arrows are overlaid to illustrate how the data are changing with time.

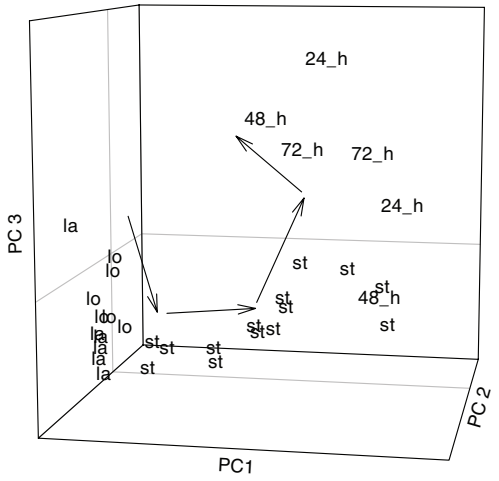

Fig. 4 *PCA plot of E. coli data where la = lag, lo = log, st = stationary. Other points are labelled with their time key, so 24_h = 24 h, 48 h = 48_h and 72_h = 72 h. The arrows have been overlaid to illustrate the change with time.*

4 RBF networks

RBFs provide an alternative approach to neural networks. The ability of RBF networks as empirical function estimators lies in these nonlinear functions being incorporated into the hidden layer of the network (see Fig. 5). Thus, the approximation problem is linear for such an approach.

RBFs are often spherical in form, being a function of the distance from the point to some constant vector, as shown in (2):

$$G(\boldsymbol{x}, \boldsymbol{c}) = F(||\boldsymbol{x} - \boldsymbol{c}||) \tag{2}$$

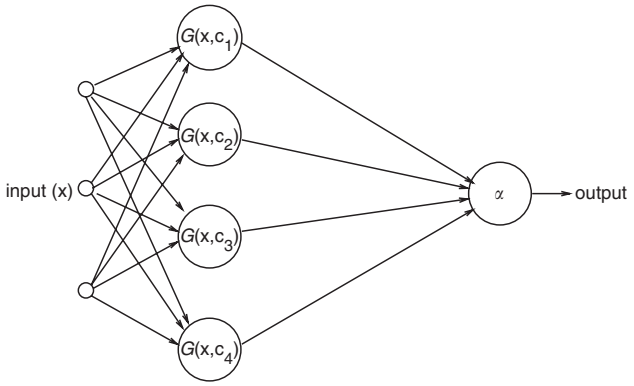

Fig. 5 *Diagram of double-layer neural network with RBFs*

Here, and for the rest of this study, $\|.\|$ is the Euclidean norm.

In the context of RBFs this constant vector is called a centre. There are a number of methods for deciding those points in the input space to use as centres. These include a regular grid of points or, in the case of the work presented here, they may be chosen from the training set.

The support vector algorithm detailed in [13] is used here. This yields a cost function of accuracy against network complexity that can be minimised using quadratic programming. In this work we consider Gaussian functions of the form:

$$G(\boldsymbol{x}, \boldsymbol{c}) = \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{c}||}{4\sigma^2}\right) \tag{3}$$

where $\boldsymbol{c}$ is the centre of the function. Hence, it can be considered as a function of a single variable, $\boldsymbol{x}$. Due to the effect the parameter σ has on the value of (3), we refer to it as the width parameter. In the context of Gaussian functions the centre $\boldsymbol{c}$ can be thought of as the mean of the function. Figure 6 shows the surface defined by $z = G((x, y), \mathbf{0})$.

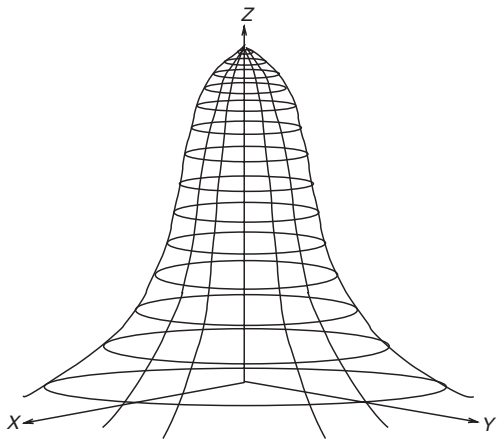

Fig. 6 *Graph of 2-D Gaussian with its centre at the origin*

The support vector algorithm considers a RBF network with one hidden layer. Each basis function in this hidden layer uses a training data point as a centre. Hence, initially there are as many basis functions as training points. However, the result of optimisation [13] is that many of the weights are estimated to be zero, reducing the number of basis functions. Hence, the number of basis functions are controlled.

There are many other RBFs that have been considered in the literature such as inverse polynomials and B-splines [8]. We consider Gaussian functions because they are simple to implement and, to a certain extent, we can visualise the way that this method operates. Furthermore, in the context of mass spectrometry analysis, the network is only concerned with the distance between points rather than what each component of a vector represents. This is advantageous because the mass spectrometry data are of a high dimension, making parameter estimation for a linear network an under-determined problem; whereas it is well-determined for a RBF network. It is thus impossible to ascertain which molecular masses or compounds are important for sample identification from the network parameters using this approach.

Looking again at Fig. 5, our network represents a discriminatory function estimator, $F(\boldsymbol{x})$ [14] of the form:

$$F(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i G(\boldsymbol{x}, \boldsymbol{c}_i) \tag{4}$$

where the $\boldsymbol{c}_i$'s are the centres, and the α_i's are the weights in the output layer, which are found during training.

The RBF networks considered here were binary prediction classifiers. One class is assigned the label '1', and the other the label '−1'. As the output of a network is a real number, this was achieved by using a three-stage activation function. Thresholds for classification were set such that an output between −0.5 and 0.5 would be classified as 'unknown'. This activation function is given in (5) and illustrated in Fig. 7.

$$\theta(t) = \begin{cases} -1 & t < -0.5 \\ 0 \quad -0.5 < & t \leq 0.5 \\ 1 & t \geq 0.5 \end{cases} \tag{5}$$

We feel this is a much better test of the accuracy of the network because it does not force the solution into two states; we need to identify an unknown bacterial species and we want to be sure that the identification is correct. The unknown region allows for the possibility that:

1) The sample is of an unknown species of bacteria.

2) The sample is of *S. aureus*, but is sufficiently different for it to be incomparable with the training set.

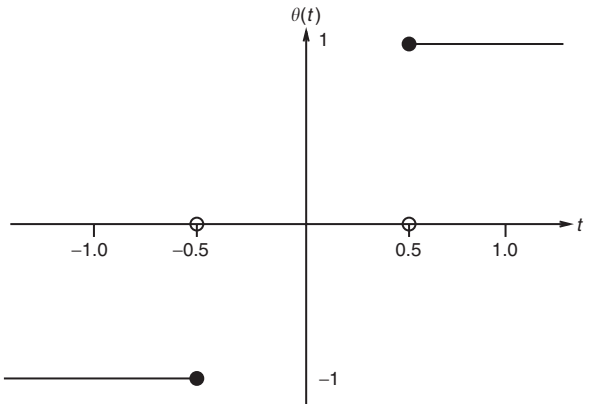

Fig. 7 *Three-stage activation function*

The networks discussed below employing Gaussian RBFs were implemented in Matlab version 6. The scripts were custom written where the alternative was to use the Matlab RBF toolbox. The decision to use our own code was based upon the control this gave, and the ease with which the data could be explored.

5 Network training regime

The two *S. aureus* experiments yielded two sets of data files. We call these data sets A and B to aid exposition. Set A contains 28 samples and set B contains 40. In each set there were equal numbers of MRSA and NCTC samples. The networks used here were trained using the support vector method. This is a batch training regime.

The *E. coli* data set consists of 32 data points. They were organised in duplicate pairs, in the sense that each sample was split into two, and placed in separate vials for sampling. The data were split up into four classes; the three main phases discussed above and a fourth phase called 'later'. The 'later' class consists of the samples taken at 24, 48 and 72 h. We introduce this fourth stage because these samples represent bacteria much further along the time scale than the other samples.

The *E. coli* data set consisted of six lag points, six log points, 14 stationary points and six 'later's'. Hence, it was necessary to extend the binary classifier. The simplest solution was to train a network for each class, each network recognising one class as being 'positive' and the others as 'negative'. This therefore required four networks in total, but only a single element output. It also permits independent training and so gives better performance of unscaled abundance data.

The four outputs were combined into a 4-D vector. Hence, each class required a positive result in the corresponding component of the output and a negative result in the other components.

Here we 'force' a classification, using the sign of the output, rather than setting thresholds. In the event of a zero output, we classify this as 'unknown'. Hence, the predicted result is the most positive component of the output vector. We feel that this is acceptable as the four classes are separated temporally and the classifications are based on 'eye' inspection of the CellFacts ouput.

5.1 Selection of training and validation data

The *S. aureus* data were collected over a full growth cycle so the data sets not only contained information on structural differences but also metabolic information [2]. To illustrate this we split the data sets into two halves, one containing early growth stage samples and the other containing later stages of growth.

It must be emphasised that, as we can train and validate with two separate sets of samples produced on different days, we can truly estimate the accuracy and the robustness of the resulting networks.

We had only one data set for the *E. coli* growth phase part of the work. However, as discussed above, the points are paired in duplicate samples. Hence, we produced two data sets by splitting up the duplicate samples. We call these two sets X and Y.

5.2 Setting the width parameter

A problem yet to be satisfactorily solved with RBF networks is setting the width parameter of the RBFs. A balance must be met where the parameter is sufficiently large to represent the distribution of the class, but also small enough so that points from another class are not included.

Many techniques have been suggested for choosing σ (see [15] and [16]) such as the average distance between training points. Here the accuracy of the networks was optimised by trial and error. The width parameter was increased in large increments until the accuracy began to drop, then the interval defined by the last two values was explored using finer steps. Although simple, this procedure was surprisingly quick, requiring very few iterations.

6 Results

We now present the results of this work in the form of confusion matrices. Each column is labelled with the true class of each sample, and the row gives the prediction. Thus, for example, in Table 2, three MRSA samples could not be identified as either MRSA or NCTC.

The accuracy is defined as the ratio of correct predictions (from 50% cross-validation) to the total number of test samples.

6.1 S. aureus

For the *S. aureus* classification problem, all of the training vectors were selected as centres, probably due to the inherent nonlinearity of the problem.

First, a network was produced using the first half of data set A, and validated with the second half. The results are shown in Table 1.

Table 1: Results of training with first half of data set A and validating with second half (the accuracy is zero)

Predicted	True classification	
	MRSA(7)	NCTC(7)
MRSA(1)	0	1
NCTC(0)	0	0
Unknown(13)	7	6

The prediction rate is very poor, and similar results were obtained from data set B. However, as discussed above, the data vary with the growth phase. This meant that the network was trained to recognise early phase data, and so could not cope with later phase data.

To explore the growth phase information contained in the data the next regime employed involved training with the first half of data set A and validating with the corresponding portion of set B and *vice versa*. The results are shown in Tables 2 and 3. These results are much

Table 2: Results of training with first half of data set A and validating with first half of data set B (the accuracy is 0.75)

Predicted	True classification	
	MRSA(10)	NCTC(10)
MRSA(7)	7	0
NCTC(8)	0	8
Unknown(5)	3	2

Table 3: Results of training with first half of data set B and validating with first half of data set A (the accuracy is 1.0)

Predicted	True classification	
	MRSA(10)	NCTC(10)
MRSA (10)	10	0
NCTC(10)	0	10
Unknown(0)	0	0

improved, especially when it can be seen that good validation was achieved against a different data set.

Finally, we extended the approach to train over an entire set. We used data set B for this test as it had the most training vectors. The result is shown in Table 4. Here we have 100% accuracy.

Table 4: Results of training with data set B (40 examples) and validating with set A (28 examples) (the accuracy is 1.0)

Predicted	True classification	
	MRSA(14)	NCTC(14)
MRSA(14)	14	0
NCTC(14)	0	14
Unknown(0)	0	0

6.2 *E. coli*

For the *E. coli* data sets 65–70% of the data were incorporated as centres in the neural network. Tables 5 and 6 detail the results using the two data sets. We do not have 100% accuracy here; in fact we have accuracies of 68.75 and 81.25% respectively.

Table 5: Results of training with set X and validating with set Y

Predicted	True classification			
	Lag(3)	Log(3)	Stationary(7)	Later(3)
Lag(3)	1	2	0	0
Log(3)	2	1	0	0
Stationary(7)	0	0	7	1
Later(2)	0	0	0	2

Table 6: Results of training with set Y and validating with set X

Predicted	True classification			
	Lag(3)	Log(3)	Stationary(7)	Later(3)
Lag	2	2	0	0
Log	1	1	0	0
Stationary	0	0	7	0
Later	0	0	0	3

It should be noted that, on the occasions when the prediction was incorrect, it was only shifted by one growth stage with respect to time. As the CellFacts classifications were (to a certain extent) subjective, the boundaries of each stage are undefined, i.e. transition between growth phases is blurred.

7 Conclusions

The headspace of certain microbial samples was analysed using a mass spectrometer. Neural networks were trained successfully to identify the class of unknown samples.

The results presented here demonstrate the application of RBF networks to analysing mass spectrometer data. By examining the principal components plots in Fig. 3 it also can be seen that, in the *S. aureus* case, RBF networks can cope satisfactorily with a nonlinear classification problem.

In this work, very little data were available. It seems, however, that theses were still sufficient enough to train accurate neural networks. Unfortunately, the small validation sets mean that we only have an estimate of the accuracy of the networks.

It may be possible, with further work, to improve classification rates by considering strain and growth phase information together. As discussed above, in the Introduction, this would, in particular, be beneficial for the determination of the correct dosage of appropriate antibiotics.

We can conclude that the technique presented here can rapidly identify certain microbial pathogens. These experiments were performed under laboratory conditions, and so more work is required to carry out these techniques under clinical conditions. Nevertheless, the technique of identification by headspace analysis shows promise as an important tool for the rapid screening of microbial infections.

8 Acknowledgments

The authors would like to thank Agilent Technologies (Delaware) for their kind provision of the Agilent 4440 and headspace autosampler; Dr. Bob Henderson for his advice and the Engineering and Physical Sciences Research Council for sponsorship of James Yates.

9 References

1 Fitzgerald, J.R., Sturdevant, D.E., Mackie, S.M., Gill, S.R., and Musser, J.M.: 'Evolutionary genomics of *Staphylococcus aureus*: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic', *Proc. Nat. Acad. Sci. USA*, 2001, **98**, (15), pp. 8821–8826

2 Esteves de Matos, R., Mason, D.J., Dow, C.S., and Gardner, J.W.: 'Investigation of the growth characteristics of *E. coli* using headspace analysis'. Proc. Olfaction and Electronic Noses, Brighton, UK, 2000

3 Gardner, J.W., Shin, H.W., and Hines, E.L.: 'An electronic nose system to diagnose illness', *Sens. Actuators B, Chem.*, 2000, **70**, (1–3), pp. 19–24

4 Gardner, J.W., Shin, H.W., Hines, E.L., and Dow, C.S.: 'An electronic nose system for monitoring the quality of potable water', *Sens. Actuators B, Chem.*, 2000, **69**, (3), pp. 336–341

5 Bellman, R.: 'Adaptive control processes: a guided tour' (Princeton University Press, Princeton, NJ, 1961)

6 Friedman, J.H.: 'An overview of prediction learning and function approximation', in Cherkassky, V., Friedman, J.H. and Weschler, H. (Eds.) 'From statistics to neural networks: theory and pattern recognition applications' (Springer-Verlag, London, 1995)

7 Haykin, S.: 'Neural networks: a comprehensive foundation' (Prentice Hall, Upper Saddle River, NJ, 1996, 2 Edn.), Chap 6

8 Smola, A.J., Scholkopf, B., and Müller, K.: 'The connection between regularization operators and support vector kernels', *Neural Netw.*, 1998, **11**, (1–3), pp. 637–649

9 Kore, R.N., Dow, C.S., and Desai, K.M.: 'A new automated system for urine analysis: a simple, cost-effective and reliable method for distinguishing between glomerular and nonglomerular sources of haematuria', *Br. J. Urology, Int.*, 1999, **84**, (4), pp. 454–460

10 Jolliffe, I.T.: 'Principal component analysis' (Springer-Verlag, New York, 1986)

11 De Maesschalack, R., Jouan-Rimbaud, D., and Massart, D.L.: 'Tutorial: the Mahalanobis distance', *Chemometr. Intell. Lab. Syst.*, 2000, **50**, pp. 1–18

12 Chang, W.-C.: 'On using principal components before separating a mixture of two multivariate normal distributions', *Appl. Stat.*, 1983, **32**, (3), pp. 267–275

13 Vapnik, V.: 'The nature of statistical learning theory' (Springer-Verlag, New York, 1995)

14 Powell, M.J.D.: 'Radial basis functions for multivariable intropolation: a review' 'Proc. IMA Conf. on Algorithms for the Approximation of Functions and Data' (Laboratories, Oxford, UK, 1985)

15 Chakravarthy, S.V., and Ghosh, J.: 'Scale based clustering using the radial basis function network', *IEEE Trans. Neural Netw.*, 1996, **7**, (5), pp. 1250–1261

16 Oukhellou, L., and Aknin, P.: 'Hybrid training of radial basis function networks in a partitioning context of classification', *Neurocomputing*, 1999, **28**, pp. 165–175