Development of a clinical data warehouse.

Panagiotis Chountas
Vassilis Kodogiannis

Harrow School of Computer Science

Development of a Clinical Data warehouse

Panagiotis Chountas
Health Care Computing Group, School of Computer Science,
University of Westminster, Northwick Park, London, HA1 3TP, UK
E-mail: chountp@wmin.ac.uk

Vassilis Kodogiannis
Mechatronics Group, School of Computer Science,
University of Westminster, Northwick Park, London, HA1 3TP, UK
E-mail: kodogiv@wmin.ac.uk

Abstract

There is increasing worldwide awareness that bionics and artificial intelligence will play an important role in microbial analysis. An intelligent data-warehouse system consisting of an odour generation mechanism, rapid volatile delivery and recovery system, and a classifier system based on Neural Networks and Genetic Algorithms have been applied as part of a microbial analysis. The microbiological warehouse environment has, also adopted the concept of fusion of multiple classifiers dedicated to specific feature parameters. The experimental results confirm the soundness of the presented methods.

1. Introduction

The *data warehousing* approach integrates data from the operational systems into one common data source, known as the data warehouse, which is optimised for intelligent data analysis purposes [1,2]. Data warehousing technology has traditionally been used in a business context, in order to answer questions about operational or executive type events in the business of concern.

The data models employed conceptually provide a multidimensional view of data, whether implemented in relational or dedicated multidimensional DBMS, and this has proven to be successful in the business application areas. However, some application areas have a need for more complex data structures. One such area is clinical data warehousing, where clinical data about large patient populations is explored to perform clinical management and medical studies. Clinical data warehousing is a large application area in itself, and we focus on describing the requirements of this area. We will concentrate on the use of clinical data for analysis purposes

The clinical domain requires more powerful data model constructs than conventional-bussiness multidimensional approaches. Let us consider the case study where a patient has multiple symptoms at the same time. The relation between patient and symptoms is modeled as an many to many relationship. This is not easily possible using a conventional multidimensional model. Using traditional dimensions, [3] we would itemise all the possible combinations of symptoms. Itemising only the combinations actually used would still yield a very large number of dimension records. In the warehouse world, the only way to reduce dimensionality is by projection, thereby disregarding all information about the omitted dimensions. It is reported [4] that a clinical DW, average patients might have hundreds of different facts describing their current situation. There is an urgent need to be able to aggregate this massive amount of information in a useful way. The issue of pre-aggregation in connection with dimensionality reduction is a very critical when it comes to data warehouses.

More advanced classification structures are also needed, including means of managing dynamic, non-strict hierarchies, and of handling change. Continuously valued data, e.g., sensor measurements, is very common and has special demands for aggregation and computation compared to conventional business data. The number of dimensions in clinical data is often very large, generating a need for intelligent ways of dimensionally reducing the data into high-level abstractions. Medical-microbiological analysis should be supported directly by the clinical data, by integrating data mining capabilities adapted to the specific domain. Integrating mining facilities will pave the way for automating managed care; a current trend in the clinical world.

Instead of relying exclusively on the judgment and knowledge of one doctor, the treatment of specific

diseases is conducted according to well-defined protocols that specify the conditions and actions. The protocol can be viewed as a "best practice" guide for patient treatment. The clinical data warehouse should have integrated support for clinical protocols, to accommodate this important part of clinical practice.

In this sense in this paper it is proposed a protocol for automating the process microbial analysis as part of a microbiological warehouses environment. The focus in this case is to enhance the warehouse functionality by integrating, dimensionality reductions techniques and data mining capabilities

2. Relevant Work

One of the biggest burdens for humanity is infectious disease. Bacteria, viruses and protozoan parasites are still major causes of death despite the steadily increasing understanding of the mechanisms of pathogenicity and the constant effort to develop novel drugs.

In the past decade, experts have started to appreciate and mimic the sense of smell by developing the electronic nose (EN). This development spans a wide range of scientific disciplines, from chemo-physics (odour detection-classification) and medicine (disease diagnosis) to psychology and psychiatry (aroma therapy). These are only a few examples of the applications of the EN. Like the light amplification by stimulated emission of radiation (*LASER*) in its early days, the EN seems to be a tool for solving many previously unanswered questions. Sensors are the most critical elements of an EN system. Sensors are technology dependent, and their commercial availability strongly depends on economic parameters such as market, available capital, and development time.

Today, in commercial EN systems, several sensor technologies are employed (conducting polymers, metal oxides, optical fibres, and piezoelectric devices) [5,6]. Many research groups around the world are actively developing new improved gas sensors with broad sensitivities to certain classes of volatile organic compounds. Over the past few years there have been an increasing number of attempts to apply artificial olfactory diagnostics in clinical practice [7]. The diagnosis of disease states is a primary pre-requisite of successful medical treatment and as such is a high priority in any area of clinical science. Microbial infections and related causes of illness seem to be one of the more common problems encountered in the world today and are widely reported by the press, especially when so-called "*killer bugs*" or "*antibiotic-resistant*" organisms are mentioned. In many cases, infection with micro-organisms produces a change in the smell of a person, which can be especially noticeable on the breath, in the urine or the stools.

Such changes have been commonly used as an aid to diagnosis of disease and in some countries, smelling the patient or the body fluids of the patient was, and still is, an important tool in diagnosis. In 1986, *National Geographic* published an article on "*The intimate sense of smell*" in which the odour of different diseases was described and in which clinicians state that odour is important in diagnosis, especially in the emergency room. However a critical step before introducing such "smart" devices into the clinic would be the i*n vitro* static or dynamic headspace analysis of microbial volatile compounds, extracted from clinical isolates of UTI, HP and respiratory infections. A metabolite may be described as volatile if it is a gas or has a high vapour pressure under the environmental conditions in which it is liberated from a cell. Organic volatile compounds (VOCs) can affect all forms of life, from the pheromones of insects, the odours of plants, to putrefaction.

Whether chemo-messengers intraspecies or interspecies (allelochemics) [8], they form complex dynamic systems of odour mixtures which can affect species behaviour and adaptation. "Table-I" presents some microbial volatiles and their biochemical precursors.

Detecting low numbers of bacterial species in clinical samples usually involves time consuming growth in selective media and subsequent isolation and identification by appropriate diagnostic procedures. Complex volatile mixtures are released during bacterial interaction with the host tissue or media, and chromatographic techniques have been used in the past to characterise those species on their gas profiles [9].

"Table I. Generation of microbial volatiles due to metabolic reaction with specific biochemical precursors"

Bacterial species	Medium	Volatile Compound
E. coli, Klebsiella sp.	Arabinose, lactose	Ethanol
Proteus sp., Klebsiella sp, Staph. Aureus, Pseudomonas sp.	Trypticase soy broth	Isobutanol, isopentyl acetate ketones
Proteus sp	L-methionine	Dimethyl sulphide, methyl mercaptan
Proteus sp, Enterococcus sp. Klebsiella sp	Acatyleholine	Trimethylamine, ethyl acetate
Proteus sp, C. septicum	Broth (complex)	Isobutylamine, isopentylamine, Ethylamine
Proteus sp	Phenylalanine, valine leucine	Benzaldehyde, isobutyraldehyde isovelaraldehyde
P. aeruginosa	Broth (complex)	Butanol, methyl ketones, 2-heptanone

Bacterial identification is a process of iterative refinement. There is no single set of tests that can be

applied to all specimens to obtain a precise identification. This process of identification is carried varying degrees of refinement depending on the body source of the specimen, the current state of microbiological knowledge, and clinical considerations. To overcome this sort of difficulty, we are exploring the use of NN simulators, on top of the microbiological data repository

3. Data Requirements

In recent years, many NN simulators have been designed and are widely available. The current state-of-the-art NN implementations can be divided into: 1) software-based simulators for use on systems such as PC compatibles and distributed system workstations and 2) hardware implementations. Each of these alternatives has strengths and weaknesses. A variety of NN architectures and training algorithms is available in the various alternatives. Some examples are: 1) back-propagation, 2) competitive networks, 3) Kohonen networks, 4) learning vector quantisation (LVQ), 5) radial basis function networks, 6) ART, 7) simulated annealing, and 8) LM method. Among all of these methods, feed-forward networks with back-propagation training algorithms are the most widely used. The appropriateness of a specific NN architecture and training algorithm is normally a function of the task and its requirements/limitations. More specifically, this appropriateness is mainly due to the input database (or the input-output structure), the complexity of the database, and the performance measures of the network. A fully connected feed-forward NN has been used by many researchers [10] in the field of artificial olfactory systems and although is not necessarily an optimal architecture proves to be reasonably effective. Based on these observations, many commercial EN manufacturers have employed feed-forward networks as their NN architecture. Regardless of the NN architecture, it has been shown that the number of neurons is of great importance in generalisation.

As a general statement, it has been shown that an excessive number of neurons can result in drastic over-training, while too few results in under-training (insufficient training) [11]. The problem of finding an optimal architecture/ configuration is extremely difficult because each specific architecture/ configuration has a unique set of optimal parameters. The objectives of this study are to:

- Discriminate *in vitro*, between multiple bacterial clinical isolates all collected from patients diagnosed with UTI, gastrointestinal and respiratory infections and reveal hidden bacterial complex patterns by using a hybrid NN-genetic algorithm mode as part of a Microbiological data-warehouse environment.

- Combine classical NN techniques with advanced AI-based methodologies (GA's) to generate a powerful hybrid classification tool; Demonstrate the power of a GA, by which a sophisticated NN can be trained for better generalisation and classification performance;
- Adopt a soft fusion of the outputs of multiple classifiers dedicated to specific feature parameters.

On of the key issues of this approach is focusing to combine the results of the various networks to give the best estimate of the optimal result.

4. Vivo-Vitro Classification

4.1 In vitro classification of bacterial clinical isolates

The following bacterial species, as illustrated in Table II, were isolated from patients suffering from Septicaemia, Respiratory, wound and Urinary Tract infections (UTI). The above clinical isolates were recovered on Blood agar plates No.2 (Oxoid), containing 5% sterile horse-blood (Oxoid) for 16hrs following primary isolation, and successful growth the biochemical profiles of all species were identified using conventional microbiological analysis performed at Gloucestershire Public Health laboratory (UK). Each one of the bacterial species was inoculated (10^6 CFU) on blood agar No 2 (Oxoid) containing 5% horse-blood (Oxoid), urea (1mg ml^{-1}), lactose (2mg ml^{-1}), L-methionine, L-valine and L-leucine (0.5mg ml^{-1}, Sigma) adjusted at pH 7.3.

Table II " General characteristics of 13 clinical isolates"

Species	Hosp.No (GRH PHLS)	Source	Diagnosis/ condition	Biochemical characteristics
Escherichia coli	72999	Urine	UTI	ß-Galactosidase$^+$, Lysine$^+$, Methyl-red $^+$, Ferm.: Lactose$^+$, mannitol$^+$, Xylose $^+$, Indole$^+$
Citrobacter spp.	94513	Blood	Septicaemia	ß-Galactosidase$^+$, Lysine$^-$, Methyl-red$^+$, Ferm.: Lactose$^-$, mannitol$^+$, Xylose $^+$.
Enterobacter cloacae	93986	Faeces	Nosocomial UTI	Extended-spectrum- ß-lactamase, Arginine $^+$, Lysine$^-$, Inositol and Glycerol: acid $^-$.
Enterococcus faecalis	72971	Urine	UTI	Acid from: Mannitol $^+$, Lactose$^+$, Sucrose $^+$, Sorbitol$^+$, Hydrolysis of Arginine $^+$, Pyruvate$^+$.
Enterococcus spp.	93981	Urine/ faeces	Cystitis	Pyruvate $^-$, melibiose$^+$.
Klebsiella oxytoca	94022	Urine	UTI	Malonate:fermentation $^+$, Glucose:gas$^+$, Indole $^+$, Urease$^+$, KCN growth$^+$.
Klebsiella pneumoniae	167913	Sputum	Pneumonia	Glucose:gas$^+$, Indole$^-$, Urease$^+$, KCN growth $^-$.
Lactobacillus spp.	71855	Gastric Juice	Gastric cancer	Catalase $^-$, H$_2$S gas$^-$.
Proteus spp.	94488	Urine	Bladder Infection	Urease $^+$, H$_2$S gas$^+$, ß-Galactosidase$^-$, Lactose$^-$.
Proteus mirabilis	94402	Urine	UTI	Urease $^+$, Ornithine decarboxylase$^+$, Lipase $^+$.
Staphylococcus aureus	94707	Wound	Skin lesion	Catalase $^+$, Coagulase$^+$,
Pseudomonas aeruginosa	73021	Sputum	Chestinfection	Oxidase $^+$, Gelatinase$^+$, NH$_3$: from Arginine $^+$.
Streptococcus pyogenes	94777	Throat swab	Sore throat	ß-Haemolysis$^+$, Group A, Ferm.: Lactose $^-$.

$^+$ Positive, $^-$Negative, **Ferm.**Fermentation, **GRH**Gloucestershire Royal Hospital, **PHLS**Public Health Laboratory Service, **UTI**Urinary Tract Infection.

All bacterial cultures were incubated at 37°C aerobically for approximately 12hrs except *Lactobacillus spp.* that was cultured micro-aerobically at 45°C and pH 6.0. A number of controls containing only sterile cultures were also incubated for the same period of time in order to study the difference between actual bacterial volatile patterns and "noisy" background produced by humidity, sensor aging and natural enzymatic digestion of cultural substrates.

I. Volatile delivery system

Following 12hrs of incubation at 37°C, each of the growing cultures-measured at the stationary phase-were placed into 2l polypropylene Mylar bags and inflated with carbon-activated filtered clean air (Hepavent, Whatman).
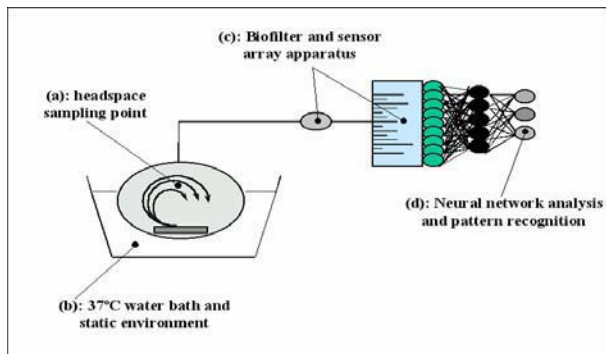

"Figure 1. Schematic representation of experimental apparatus"

Each bag was transferred into a 37°C water bath and left to equilibrate for 5min before being connected with the sensory unit through a 15cm long Teflon tubing, a hydrophobic PTFE filter (Hepavent, Whatman), to ensure a sterile less humid environment over the sensor surfaces. The sampling point was adjusted to a set height above the static headspace as illustrated in "Figure 1". A flow rate of 200ml min^{-1} was set automatically by data control software. Additionally environmental conditions at the sampling point, inside the water bath were continuously monitored in order to establish a standardised sampling protocol.

II. Bacterial pattern recognition

"Figure 2" displays a real time sensory response analysed by 5 extracted sensor features that describe sensor-volatile physicochemical interaction and pattern extraction: (a) Divergence: maximum step response, (b) Absorption: maximum rate of change of resistance), (c) Desorption: maximum negative rate of change of resistance, (d) Area under the curve and (e) Ratio Absorption/Desorption. In order to improve the bacterial classification process fourteen conducting polymers and the above 5 features generated a set of 70 sensor parameters. All sensors responses were pre-processed by using a suitable normalisation algorithm [12].
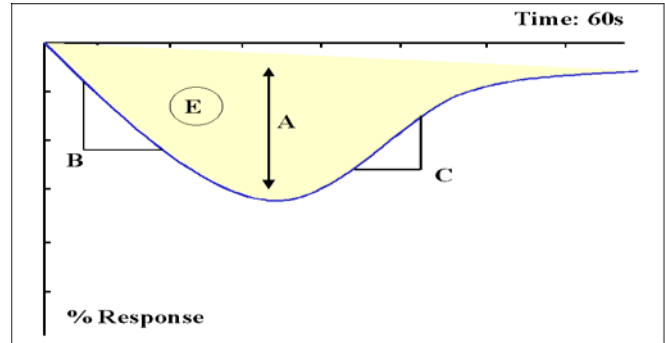
"Figure 2: Parameters measured for each sensor response"

4.2 In vitro classification of Urinary infections

UTI is a significant cause of morbidity with 3 million UTI cases each in the USA alone [13]. Thirty-one percent of nosocomial infections in medical intensive care units are attributable to UTI, and it is estimated that 20% percent of females, aged of 20 and 65 years suffer at least one episode per year. There are also links to other complicated or chronic urological disorders such as pyelonephritis, urethritis, and prostatitis[14]. Approximately 80% of uncomplicated UTI are caused by *E.coli* and 20% by enteric pathogens such as *Enterococci*, *Klebsiellae*, *Proteus sp.*, coagulase (-) Staphylococci and fungal opportunistic pathogens such as *Candida albicans* [15]. Current diagnostic techniques require 24-48 hrs to identify pathogenic species in urine midstream specimens ($\geq 10^5$c ml^{-1}) and apply antibiotic sensitivity tests. Despite the introduction of molecular tests, microscopy and culture remain the gold standard in every day clinical practice.

I. Urine samples and volatile generating kits

Forty-five 5ml urine samples (following eukaryotic cell filtering extraction) were collected from randomly selected patients admitted in Gloucestershire PHLS and inoculated into specially made centrifuge bottles (50ml, Sterilin) each containing 95% BHI broth (Oxoid), 5% serum bovine (Oxoid), 0.70mg ml^{-1} of a series of amino acids (L-Leucine, L-Alanine, L-Serine, L-Valine, L-Asparagine, L-Glutamine, L-Methionine, Sigma), 1mg ml^{-1} Urea (Sigma), 0.75mg ml^{-1} Lactose (Sigma), 0.1mg ml^{-1} Casein (Oxoid), 0.3mg ml^{-1} Acetylcholine (Sigma) to

a final volume of 20ml per VGK and incubated aerobically for 5 hrs at 37°C.

II. Flow injection analysis (FIA) of urinary volatiles

After 5 hrs of incubation to coincide with the logarithmic phase of growth, 45 VGK were placed in a 37°C water bath and directly connected with a specifically designed air-filtered sparging (bubbling) system. This consisted of Teflon tubing (Tygon), a hydrophobic biofilter (0.45μm PTFE, Whatman-Hepavent) and an activated carbon filter (Whatman) to provide clean air-flow above the urine headspace. A flow rate of 200ml min^{-1} was set automatically and environmental conditions at the sampling point were continuously monitored. The actual urine sampling time and baseline recovery per specimen was 3 min.

III. Intelligent UTI pattern recognition system

Thirty cases of UTI were identified from 45 randomly selected samples by standard microscopy and culture: 13 patients were infected with E.coli (e), 9 with Proteus sp. (p) and 8 with coagulase (-) Staphylococcus sp., (st). Two genetic training algorithms processed urine data through a parallel evolutionary succession process towards competent NN solutions. The first GA analysed patient data that had been randomly divided into a "training" group of 31 urinary samples (e: 9, p: 6, st: 5 and n: 11) and a group of 14 "unknowns" (e: 4, p: 3, st: 3 and n: 4, 31% of patient collected data).

5. Odour Recognition and Data Analysis

5.1 In Vitro Analysis

Two hundred and forty-eight bacterial patterns of 14 classes and 70 normalised sensor parameters constructed a matrix of 17,360 sensor data-items that was analysed by an intelligent system consisting of Radial Basis-Function-Networks (RBF).
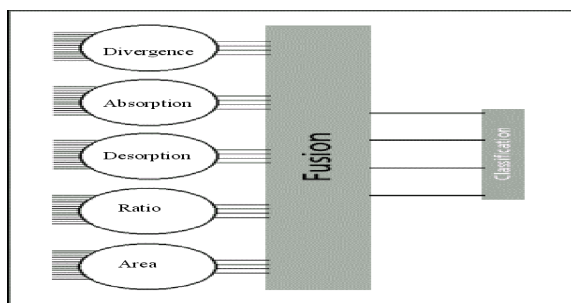

"Figure 3. Multiple classifier architecture"

Overall, the sensor data matrix was randomly divided

into a training group containing 200 bacterial patterns and a testing one of 48 random "unknown" samples.

Recently, the concept of combining multiple networks has been actively exploited for developing highly reliable neural network systems. One of the key issues of this approach is how to combine the results of the various networks to give the best estimate of the optimal result. A straightforward approach is to decompose the problem into manageable ones for several different sub-networks and combine them via a gating network. The proposed architecture is a neural network system containing five parallel modules, one for each of the bacterial properties as shown in "Figure 3". Each network module makes a classification from a single property and their results are combined, using an averaging approach, to make an overall classification. All modules contain fourteen input nodes and four output nodes. The fourteen input nodes correspond to the fourteen sensor parameters.
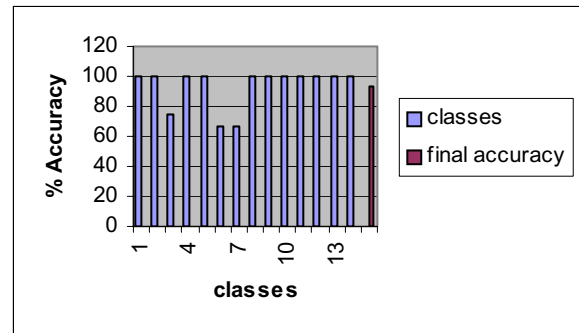
"Figure 4. Fusion analysis results"

The four output nodes are sufficient for binary representation of the fourteen classes (13 bacterial classes and the control group). Four binary digits can represent sixteen integers, so each class is assigned one of fourteen binary patterns. Two patterns remained unused. The soft combination of neural classifiers resulted in 93.75% accuracy over the testing dataset, demonstrating in this way the efficiency of this scheme in terms of accuracy and processing-time. The relevant results are illustrated in "Figure 4".

5.2 In *vivo* analysis

An evolutionary process of 5 generations (3 NNs/ generation) was carried out employing 1 crossover and a mutation rate of 0.5. Additionally the second GA performed a much broader evolutionary optimisation analysis of 100 generations. It also attempted to analyse the same amount of patient data but with a higher ratio of "unknown" proportion (42% of collected patient data) including 26 training samples (e: 8, p: 4, st: 4, n: 10) and

19 "unknown" UTI (e:5, p:5, st:4, n:5).

A population of 600 NNs was evolved using an immigration mode, 2 crossovers and a mutation rate of 0.7 towards the "fittest" NN solution. Both "genetically" selected sensor parameters were also used to perform PCA and DFA-cv. PCA accomplished non-parametrically a significant dimension reduction by minimising minor UTI data variations so that information could be depicted on a few two-dimensional principal component score plots. Two parallel evolutionary algorithms selected 2 NN solutions.

The first was a 3-layer (28-12-4) back-propagation NN that used an adaptive learning rate, a momentum of 0.42, an input pattern noise of 0.03 and achieved a 98% prediction rate. Thirteen out of 14 "unknown" UTI samples were identified correctly with a prediction output confidence ranging from 0.75 to 1.01. The intelligent system failed to characterise only one urine sample previously diagnosed with *E.coli* infection.
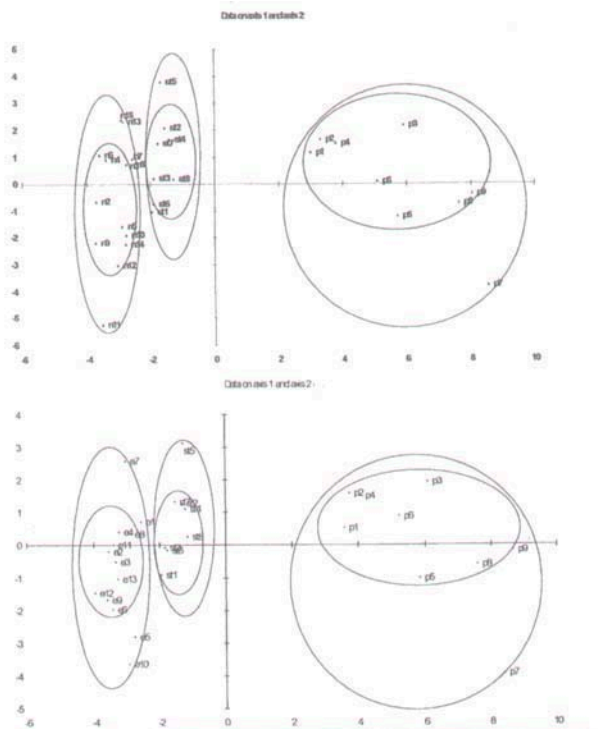
"Figure 5. Extraction of "genetically" selected sensor parameters and two-dimensional representations of PCA clustering between: a. normal urine (n), Proteus sp. (p) and Staphylococcus sp. (st) and b: E.coli (e), Proteus sp. (p) and Staphylococcus sp. (st). (Inner and outer circles divide most closely linearly discriminated patterns from the most drifted ones, respectively)."

However, this single pattern confusion was limited to the case of distinguishing between *E.coli* infection and normal urine. Both their prediction confidence outputs were very close-0.37 for *E.coli* and 0.43 for normal urine-but below a 0.5 test tolerance limit.
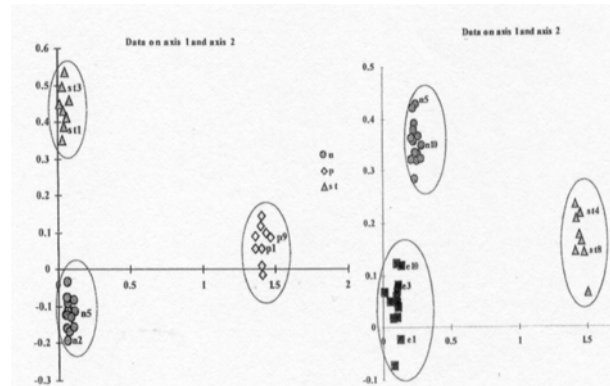
"Figure 6. DFA and 3-group separation between: a. normal urine (n), Proteus sp. (p) and Staphylococcus sp. (st) and b. normal urine, E.coli (e) and Staphylococcus sp."

Twenty-eight "genetically" selected parameters performed PCA and DFA, which displayed two graphical cluster separations between *Proteus sp.*, *Staphylococcus sp.* UTI and normal samples. Cross-validation reclassified correctly 6 "unknown" patient samples "Figs 5a & 6a". Furthermore by extracting all "genetically" selected sensor parameters that had been previously used as input neurones it was possible to reveal hidden non-linear patterns characteristic of each UTI group. Furthermore the second 3-layer NN (22-15-4) achieved a 95% prediction rate and recognised 18 out 19 "unknown" UTI cases. Only one normal patient sample had been mistaken for E.coli infection.

A two-dimensional discrimination plot between 3 of the tested UTI groups (e, st, p) was produced by PCA. DFA also separated patient samples infected with E.coli, Staphylococcus sp. and normal urine samples. Cross validation recognised 7 "unknown" UTI cases (Figures 5b & 6b).

7. Conclusions

We have shown that DW technology faces new challenges from the area of microbiological data repositories. The integration of clinical protocols in the CDW is important to allow for follow-up on the treatment of patients. Support for medical research, e.g., via data mining facilities, will enable the clinical community to perform their research much more efficiently than is possible today.

Microbiological data warehousing provides prospect for large repositories research that will also have applications in areas beyond the clinical-biological world.

There is need for innovative inexpensive tests to be developed for early diagnosis of infectious diseases and control of antibiotic resistance. The recent use of GC-MS or MS methods accompanied by NN and multivariate analysis although are considered very sensitive, they need highly skilled personnel and are characterised by increasing capital cost. Intelligent gas sensor technology has been applied in several research areas, including biomedicine. Many research groups around the world are actively developing new improved gas sensors with broad sensitivities to certain classes of volatile organic compounds. As these sensors become commercially viable, the EN might well achieve higher levels of acceptance in medical applications.

The present system resulted in the delivery of bacterial odours in the form of repetitive 'sniffs', and achieved higher control by keeping the sampling point, the headspace and liquid volumes constant. Additionally there was continuous monitoring of environmental conditions at the sampling point. There are several advantages in the application of NN models as opposed to other statistical techniques. Their ability to generalise is particularly useful since rough data is often noisy due to some sensor drift. Selecting and constructing the right learning data (input) is crucial in pattern recognition methods. Each class must be composed of representative and reproducible samples. The quantity of these samples does not increase the discrimination confidence instead it is the "quality" of representation carried in each input sample that determines pattern recognition performance.

The applied GA-NN technique achieved a high prediction rate and enabled the parallel use of multivariate techniques too, showing a degree of correlation among genetically selected input parameters. The present work proposes a novel application of GA-NN in combination with multivariate techniques in bacterial class discrimination. However, the use of multiple NN fusion is a challenging and more promising approach. The adopted parallel architecture reduces the dimensionality of the network search space thus increasing both computational efficiency and the probability that optimal network parameters will be found within the search space. Future work will investigate the integration of neuro-fuzzy algorithm to the multiple classifier scheme employed however with a more accurate fusion decision criterion, such as the fuzzy integral.

References

1. Widom J., Research Problems in Data Warehousing in *Proceedings of CIKM 1995*, pp. 25–30.
2. Widom, J., Special Issue on Materialized Views and Data Warehousing. *IEEE Data Engineering Bulletin*, 18(2), 1995.
3. Kimball, R., *The Data-Warehouse Toolkit*. Wiley Computer Publishing, 1996.
4. Chountas, P., Petrounias I., Atanassov K., Kodogiannis, V. El-Darzi E., Representation & Querying of Temporal Conflict" FQAS 2002, DBLP, LNAI, No. 2522, pp.112-123, Springer-Verlag
5. Gopel, W., Chemical Imaging: I. Concepts and Visions for Electronic and Bioelectronic Noses. *Sensors & Actuators B*, 52, 125-142 (1998).
6. Gardner, J.W., Barlett P.N., A brief history of electronic noses. *Sensors & Actuators B,* 18, **1-3**, 211-220 (1994).
7. Gardner, J.W., Shin, H.W., Hines, E.L., An electronic nose system to diagnose illness. *Sensors & Actuators B,* 70, 19-24 (2000).
8. Davies, T., Hayward, N.J., Volatile products from acetylcholine are markers in the rapid urine test using headspace gas liquid chromatography. *J. Chromatography*, 307, 111-121 (1984).
9. Cox, C.D., Parker, J., Use of 2-aminoacetophenone production in identification of Pseudomonas aeruginosa. *J. of Clinical Microbiology*, 9, 479-484 (1979).
10. Kermani, B.G, Schiffman, S.S, Nagle, T, Using Neural Networks and Genetic Algorithms to Enhance Performance in an Electronic Nose, *IEEE Transactions on Biomedical Engineering*, 46, **4**, 429-439 (1999).
11. Kodogiannis, V. Comparison of advanced learning algorithms for short-term load forecasting, *Journal of Intelligent & Fuzzy Systems*, 8, **4**, 243-259 (2000)
12. A.K. Pavlou, V.S. Kodogiannis, A.P.F. Turner, "Intelligent classification of bacterial clinical isolates *in vitro*, using electronic noses", *Int. Conf. on Neural Networks and Expert Systems in Medicine and HealthCare, Greece,* pp. 231-237, 2001.
13. Schaechter M, Medoff G, Eisenstein BI, eds. *Mechanisms of microbial disease*, 2nd edn. Baltimore: Williams & Wilkins, 1993.
14. Lipsky BA. "Prostatitis and urinary tract infection in men: what's new; what's true?" *Am J Med,* Vol. 106, pp. 327-34, 1999.
15. Honkinen O, Lehtonen OP, Ruuskanen O, Huovinen P, Mertsola J. "Cohort study of bacterial species causing urinary tract infection and urinary tract abnormalities in children", *BMJ,* Vol. 318, pp. 770-71, 1999

IEEE COMPUTER SOCIETY