Application of Simulated Annealing Fuzzy Model Tuning to Umbilical Cord Acid-Base Interpretation

Jonathan M. Garibaldi and Emmanuel C. Ifeachor

***Abstract*— Fuzzy logic and fuzzy set theory provide an important framework for representing and managing imprecision and uncertainty in medical expert systems, but the need remains to optimize such systems to enhance performance. This paper presents a general technique for optimizing fuzzy models in fuzzy expert systems (FES's) by simulated annealing (SA) and N-dimensional hill climbing simplex method. The application of the technique to a FES for the interpretation of the acid-base balance of blood in the umbilical cord of newborn infants is presented. The Spearman rank order correlation statistic was used to assess and to compare the performance of a commercially available crisp expert system, an initial FES, and a tuned FES with experienced clinicians. Results showed that without tuning, the performance of the crisp system was significantly better (correlation of 0.80) than the FES (correlation of 0.67). The performance of the tuned FES was better than the crisp system and effectively indistinguishable from the clinicians (correlation of 0.93) on training data and was the best of the expert systems on validation data. Unlike most applications of fuzzy logic where all fuzzy sets have normalized heights of unity, in this application it was found that a reduction in the height of some fuzzy sets was effective in enhancing performance. This suggests that the height of fuzzy sets may be a generally useful parameter in tuning FESs.**

***Index Terms*— Acid-base balance, fuzzy modeling, neonatal outcome, simplex method, simulated annealing.**

I. INTRODUCTION

ONE challenge in medical expert systems is the problem of imprecision and uncertainty in both data and knowledge [1]. In practice, it is important to develop techniques for handling such imprecision and uncertainty to enhance the robustness and performance of medical expert systems. Fuzzy logic and fuzzy set theory [2] provide a good framework for managing uncertainty and imprecision in medicine [3]–[5] and have been successfully applied to a number of areas [6], [4], [7].

The successful development of a fuzzy model for a particular application domain is a complex multistep process in which the designer is faced with a large number of alternative implementation strategies. The principal alternatives are in the selection of:

- inference methodology;
- linguistic variables and fuzzy terms;
- rule set;
- fuzzy operators;
- membership functions.

The effect of a fuzzy system is defined by the set of vectors comprising the fuzzy output variables that are obtained from a set of vectors representing input variables. Even the simplest modification to the fuzzy system such as the use of linear rather than nonlinear membership functions may alter the input–output mapping of the fuzzy model. Thus, the behavior of a fuzzy system is governed by a combination of all these design choices. In a given application there is the additional process of defuzzification to map the output fuzzy sets to the real world. Unfortunately, there is currently very little theoretical guidance as to which of the design choices are "correct" for a particular domain.

A number of approaches have been proposed for the development, tuning, and optimization of fuzzy models. Many of these have been based on the integration of neural networks with fuzzy logic [8], [9] or hybrid neuro-fuzzy clustering methods such as Fuzzy-ARTMAP [10] or ANFIS [11]. More general algorithmic approaches have included the Σ-PAFIO algorithm for model optimization [12], [13], a framework for synthesizing fuzzy rules [14], and *genetic algorithm*—based optimization methods [15]. The well known *simulated annealing* (SA) algorithm [16] has been applied to the tuning of fuzzy neurons [17], the combinatorial optimization of fuzzy partitioning [18] and for multivariable optimization of a fuzzy relational model-based controller [19]. In this paper, *SA* with an adaptation for continuous minimization by the simplex method [20] is applied in a novel manner to tuning of fuzzy models. This general purpose algorithm can be used for both combinatorial and continuous parameter optimization and, thus, is suitable for both structure identification and parameter estimation of models. The optimization technique is described and then its use is illustrated with application to the automated tuning of a fuzzy expert system (FES).

A successful crisp expert system for the interpretation of the chemical acid-base balance of blood taken from the umbilical cord of newborn infants had previously been developed [21], which incorporated the knowledge of several expert clinicians. This crisp system was extended to incorporate explicit handling of imprecision in the input data and uncertainty in the embedded knowledge. This crisp expert system formed the basis of the FES—the existent crisp rule set was used to derive the initial fuzzy rule set and to guide the initial choice of location of membership function for each fuzzy term. This medical domain is characterized by the lack of an external "gold-standard" criteria against which to validate the output of

Manuscript received January 10, 1997; revised May 13, 1998. This work was supported by the EPSRC; crisp expert system software is available as *Expert DataCare* Chiron Diagnostics Ltd., Halstead, Essex, U.K.

The authors are members of the School of Electronic Communication and Electrical Engineering, University of Plymouth, Drake Circus, Plymouth, PL4 8AA U.K.

Publisher Item Identifier S 1063-6706(99)01369-7.

1063–6706/99$10.00 © 1999 IEEE

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

the crisp or FES's. Hence, a validation study was undertaken in which six expert clinicians reviewed 50 difficult abnormal cases to rank them in order from "worst" to "best" and these rankings were compared against those obtained for the expert systems.

It was found that the initial FES performed worse than the crisp system and, thus, it was decided to use the clinicians' ranking of the 50 cases as training data to automatically tune the performance of the fuzzy system using the SA algorithm described. Normalization of each linguistic term has been advocated as a necessary condition in fuzzy models [13] with some theoretical justification. However, there is little evidence that human experts use logically sound reasoning in their decision making and, thus, theoretical soundness may be less important in expert system applications of fuzzy modeling than in fuzzy control applications. In this application, the requirement for normality was relaxed and the height of fuzzy terms was included as a tunable parameter within the modeling process. Using this method the performance of the FES was improved to attain a level comparable to the clinicians. Finally, an external criterion known as the *Apgar score* was used to validate the performance of the pre- and post-tuning FES.

II. Development of an Optimization Technique

A. Simulated Annealing

The formulation of a fuzzy model appropriate for a particular domain involves a large number of choices, many of which have no theoretical basis. In general, the formulation of an optimal fuzzy model in terms of its performance at a given task in the particular domain is a problem of N-dimensional nonlinear optimization in which N is very large even for the most trivial of fuzzy systems.

The *SA* algorithm is a general-purpose algorithm for performing approximate optimization in large dimensional problems [16]. It is generally useful for combinatorial optimization problems and/or problems where derivatives of the cost function being optimized are not available. The SA algorithm views the cost function $f(x)$ being minimized as equivalent to the energy state of a physical system [22] and the process of reaching equilibrium is equivalent to repeatedly accepting or rejecting changes in energy state from i to j according to the probability

$$\boldsymbol{Prob}(\text{accept change to state } j \text{ from } i = \begin{cases} \exp\left(\dfrac{f(i)-f(j)}{T}\right), & \text{if } f(j) > f(i) \\ 1, & \text{if } f(j) \le f(i). \end{cases} \quad (1)$$

The algorithm depends on two control parameters: the temperature T and the number of state changes L considered at each temperature. An initial state is generated and then L state changes are considered according to the criterion in (1). The temperature T is then reduced according to a *cooling schedule* and the process is repeated until an appropriate stopping criterion is reached. Typical stopping criterion include; when the cost function is minimized to a certain value when the changes in energy being considered are sufficiently small or when the temperature is sufficiently low. Theoretical convergence properties and overall efficiency of the algorithm depend on the cooling schedule chosen—it has been shown that the algorithm is guaranteed to optimize $f(x)$ in infinite time [23]. In practice appropriate values of T, L and the cooling schedule can be determined experimentally to return near-optimal solutions. The algorithm is detailed in Appendix A.

B. Continuous Minimization by the Simplex Method of Simulated Annealing

The SA algorithm implements a local "valley-descent" algorithm in which downhill changes are always accepted and uphill changes are accepted initially, but with decreasing probability as the algorithm progresses. The technique has been applied most often to discrete combinatorial optimization problems, but it has also been applied to problems with continuous variables. With continuous variables it is necessary to minimize a function $f(\boldsymbol{y})$ where $\boldsymbol{y}$ is an n-dimensional vector in $\Re^n$ with components $(y_1, y_2, \cdots, y_n)$. The change in state is taken as a move in hyperspace to $\boldsymbol{y}+\Delta\boldsymbol{y}$ so that the change in energy is $f(\boldsymbol{y}+\Delta\boldsymbol{y}) - f(\boldsymbol{y})$. A number of methods by which to choose the direction and length of the step $\Delta\boldsymbol{y}$ have been proposed [24]–[26], including a modification of the downhill simplex method of Nelder and Mead [27].

A simplex is a geometrical figure of $n+1$ vertices in n-dimensional space, for example a triangle in two-dimensional space or a tetrahedron in three-dimensional space. The simplex is initialized with a set of $n+1$ points, and the function is evaluated at each of these points. The simplex then makes one of the following steps:

1) a reflection of the vertex with the highest function evaluation through the opposite face to a point that conserves the "volume" of the simplex;
2) a reflection of the highest vertex through the opposite face to a point that increases the "volume" of the simplex;
3) a contraction of the highest vertex toward the opposite face of the simplex;
4) a contraction along all faces toward the lowest vertex of the simplex.

The simplex algorithm specifies the series of these steps to take in such a way that the simplex converges to a (local) minimum. In the SA adaptation of the simplex method [20] a positive, logarithmically distributed random variable proportional to the temperature T is added to the function evaluation at each vertex of the simplex when determining the highest and lowest points so that

$$g(\boldsymbol{y}) = \boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{T}\ \ln(\mathrm{random}(\boldsymbol{0},\boldsymbol{1})) \quad (2)$$

and a similar value is subtracted from the function evaluation of each new point that is tried as a replacement point. Thus, a downhill step is always accepted and an uphill step is accepted with probability $\propto \exp(-\text{uphill step-size}/T)$. In the limit $T \to 0$, this algorithm reduces to the downhill simplex and converges to a local minimum. As with any form of the SA algorithm the choice of initial temperature, step length (simplex size), and cooling schedule will determine whether

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

the global minimum will be found successfully and these parameters will usually be found by trial and error. The algorithm is detailed in Appendix B.

A problem occurs with the simplex algorithm when $\boldsymbol{y}$ is bounded, especially when the global function minimum is located on or very near to the boundary (as is often the case in continuous variables used in fuzzy models). The standard procedure when not using the simplex adaptation [25] is to simply reject any step that generates $\boldsymbol{y}$ outside its bounds. However, if the simplex adaptation is used and vertices that lie outside the bounds are rejected, then a global minimum that lies on a boundary will never be enclosed by the simplex and, thus, will not be successfully located. To overcome this problem, if a $\boldsymbol{y}$ is generated with at least one y_i outside a bound y_i' a penalty term can be added, so that

$$g(\boldsymbol{y}) = g(\boldsymbol{y}') + \gamma\left(\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{y}_i')^2\right) \tag{3}$$

each y_i outside a bound y_i' where $g(\boldsymbol{y}')$ is the function defined by (2) evaluated at the boundary.

C. Simulated Annealing Optimization of a Fuzzy Model

A fuzzy model may be viewed as a N-dimensional vector characterizing a FES with a set of m discrete variables and n continuous variables. Thus

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{f}(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m, \boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n) \tag{4}$$

where each x_i is a discrete variable and each y_j is a continuous variable, as above. Discrete variables represent structural choices such as which rules from the total set of possible fuzzy rules to use or which of the set of operator families to implement; continuous variables represent parameters such as modal values or spreads of membership values or the tunable parameter β in the Yager class of operators [28].

If a cost function can be found to evaluate the performance of a fuzzy model with each set of x_i and y_j then the SA algorithm can be applied to optimize the model. If no continuous variables are used, then the discrete SA algorithm (Section II-A) may be used alone and if no discrete variables are used then the continuous SA algorithm (Section II-B) may be used alone. However, if both discrete and continuous parameters are to be otpimized simultaneously, then a slightly more sophisticated SA must be used. In this case, a decision must be made to either generate a new state through a change in the discrete variables or via a simplex move in the continuous variables. This decision can be made simply as a random function of m and n (the number of discrete and continuous variables, respectively) or through a more sophisticated criteria dependent on $\Delta f(\Delta \boldsymbol{x}, \boldsymbol{y})$ and $\Delta f(\boldsymbol{x}, \Delta \boldsymbol{y})$; that is, the relative changes in the cost function that result from changes in only the discrete variables or only the continuous variables. The algorithm is detailed in Appendix C.

The selection of *stopping criterion* is a difficult problem. If the optimal value of the cost function f_{opt} is known *a priori*, and given an infinite amount of time such that the cooling schedule can be sufficiently slow, then the algorithm can be terminated when $f - f_{\text{opt}} < \epsilon$. In practice, f_{opt} is unknown and a finite cooling schedule must obviously be used. The choice of reasonable cooling schedule and appropriate stopping criterion is the hardest part of successfully implementing the tuning algorithm and the solution will be problem dependent. Theoretical issues have been investigated [23] and practical strategies have been developed [23], [26], [20]. For fuzzy model tuning, a relatively simple strategy may be implemented such as:

- reduce T to ωT where $0 < \omega < 1$, terminate when $T \leq T_0$ where T_0 has previously been chosen as the minimum temperature;
- reduce T to ωT, terminate when $\Delta f < \epsilon$, i.e., when the improvement obtained in f in an iteration is sufficiently small.

D. Obtaining a Cost Function

In general, *any* cost function can be used: it does not need to be smooth or even continuous, but it must be bounded [26]. In practice a measure of mean-square error between actual and some desired target output would be appropriate to tuning a fuzzy model. Usually, the output of a fuzzy system will be arbitrarily scaled, making an absolute comparison between the fuzzy system output and target output difficult. In such a situation, a statistic that ignores the absolute values, but takes into account the ordering of results may be appropriate. Spearman rank-order correlation can be used to determine the degree of association between two sets of rank-ordered data. It is defined as

$$\rho_s = \frac{\sum_{i=1}^{K} (r_i - \overline{r})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{K} (r_i - \overline{r})^2 \sum_{i=1}^{K} (s_i - \overline{s})^2}} \tag{5}$$

where r_i and s_i are the two sets of rankings being compared and K is the number of ranks [29]. If there are no ties in the rankings then (5) can be simplified to give

$$\rho_s = 1 - \frac{6\sum_{i=1}^{K} (r_i - s_i)^2}{K(K^2 - 1)}. \tag{6}$$

The ρ_s statistic will be $+1$ if the two rankings are identical (complete agreement) and -1 if the rankings are opposite (complete disagreement). The significance of the statistic can easily be obtained by

$$z = \rho_s\sqrt{K - 1}, \quad \text{with} \quad df = K - 2$$

if K is larger than about 20. If r_i are the desired rankings that have been determined for a particular set of input parameters to a FES and s_i are the actual rankings obtained from the fuzzy system, then $f = 1 - \rho_s$ is a suitable cost function to be minimized by SA. Note that this is effectively the same as minimizing the mean-square error between the desired rankings and the obtained rankings.

In practice, a target set of rankings may be obtained in a number of ways. In many cases, an objective *correct* ordering of a set of results may be available, but for situations where

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

such a verified ordering is not available, a target set may be obtained by seeking the opinions of as many human domain experts as is feasible. If k different expert opinions are obtained, then the interexpert agreement can be calculated from their rankings to ensure that there is an acceptable body of opinion. The pairwise average rankings may thus be calculated to be used as the target set for SA of the FES or the individual rankings may be kept separate to allow the fuzzy system to be trained to certain experts. Note that if the average correlation between all pairs is denoted $\text{ave}(\rho_s)$, then $\text{ave}(\rho_s)$ can only range from $-1/(k-1)$ to $+1$, as it is not possible for k experts to all totally disagree with each other [29]. In this case, the significance of $\text{ave}(\rho_s)$ must be tested by

$$\chi^2 = (K-1)((k-1)\,\text{ave}(\rho_s)+1), \quad \text{with} \quad df = K-1.$$

III. Application to Umbilical Acid-Base Interpretation

A. Medical Background

In many areas of medicine such as obstetrics, advances in technology have led to a bewildering choice of equipment for the doctor and other healthcare professionals. Much of the available technology produces results that require considerable time and expertise to interpret correctly, but the routine day to day care of patients is carried out largely by junior clinical staff without the necessary experience to utilize the technology effectively [30]. The need for expert systems in medicine has long been established [31]. They allow a better match of available technology to the needs of the user and offer the potential for consistent, reliable, and objective analysis of data using captured expert knowledge and presenting results in a way that will practically aid the busy clinician.

The process of natural childbirth is a stressful experience for both mother and infant. Objective assessment of the neonatal outcome of labor is difficult and challenging. Accurate assessment of immediate neonatal condition can be used to guide individual neonatal care and to provide feedback to clinicians reviewing their practice. Current methods (such as the Apgar score) are inadequate as they are too subjective and/or they fail to distinguish damage that occurred during labor from damage that occurred before or after labor. Analysis of blood taken from the umbilical cord, clamped immediately after delivery, can be used to obtain objective information on the physiological state of the infant.

The umbilical cord vein carries blood from the placenta to the fetus and the two smaller cord arteries return blood from the fetus. The blood from the placenta has been freshly oxygenated and has a relatively high partial pressure of oxygen $(p\text{O}_2)$ and low partial pressure of carbon dioxide $(p\text{CO}_2)$. Oxygen in the blood fuels *aerobic* cell metabolism with carbon dioxide produced as *waste*. Thus, the blood returning from the fetus has a relatively low oxygen and high carbon dioxide content. Some carbon dioxide dissociates to form carbonic acid in the blood, which increases the acidity (lowers the pH). A degree of oxygen starvation for the fetus is a routine occurrence in normal labor. If oxygen supplies are too low, *anaerobic* (without oxygen) metabolism can supplement aerobic metabolism to maintain essential cell function, but this produces lactic acid as *waste*. This further acidifies the blood and can indicate serious problems for the fetus.

Samples of blood may be taken from blood vessels in the umbilical cord of the neonate immediately on delivery and a blood gas analysis machine measures the pH, partial pressure of carbon dioxide $(p\text{CO}_2)$, and partial pressure of oxygen $(p\text{O}_2)$. A parameter termed the *base deficit of extracellular fluid* (BD_{ecf}) can be derived from the pH and $p\text{CO}_2$ parameters [32]. This can distinguish the cause of a low pH between the distinct physiological conditions of *respiratory* acidosis due to a short-term accumulation of CO_2 and a *metabolic* acidosis due to lactic acid from a longer term oxygen deficiency. Analysis of the acid-base balance of arterial and venous blood from a clamped umbilical cord provides objective information on the severity and duration of any lack of oxygen during labor [33]. However, the procedure is error prone and a 25% failure rate to obtain the parameters necessary for an accurate interpretation is not uncommon [34].

B. The Crisp Expert System

A rule-based expert system to validate and interpret umbilical cord blood acid-base status has been developed in close collaboration with several experienced clinicians and has been implemented in the local hospital and a number of other hospitals in the U.K. [21]. First, the pH $p\text{CO}_2$ and $p\text{O}_2$ data from the arterial and venous samples are validated as being consistent with possibilities for real cord blood. The pH and BD_{ecf} are then examined to categorize the results into one of 54 interpretations, ranging from "normal" to "severe metabolic acidemia." Results consistent with respiratory acidosis are distinguished from those indicating metabolic acidosis and the differences between the two vessels, if available, are used to further refine the diagnosis. This system has been implemented on labor ward at the local hospital to collect data from every delivery where a database of over 12 000 results has been collected so far.

C. The Need for a Fuzzy Expert System

A number of problems were identified in the implementation of conventional crisp rules used in the initial system. The interpretation section of the crisp expert system utilized a number of rules of a form similar to

IF arterial pH $<$ 7.05 AND arterial $\text{BD}_{\text{ecf}} \geq 12$ mmol/l
THEN severe arterial metabolic acidemia.

Such rules feature sharp boundary cutoffs, which are not representative of real decision-making processes and do not employ any form of uncertainty representation in the conclusion to imply a less than certain diagnosis. There was a need for a fuzzy logic-based expert system that would offer a more realistic and acceptable interpretation.

A preliminary investigation was performed to convert the crisp expert system directly into a FES. The purposes of this study were

- to determine the feasibility of converting the existing crisp rules into a set of fuzzy rules without the necessity of additional expert knowledge elicitation sessions;

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

TABLE I
AN EXAMPLE TO JUSTIFY THE USE OF THE PROBABILISTIC CONJUNCTION OPERATOR $(a * b)$ RATHER THAN THE MINIMUM CONJUNCTION OPERATOR $\min(a, b)$

Case	pH_A	μ_1	BD_A	μ_2	pH_V	μ_3	BD_V	μ_4	$\min(\mu_1, \mu_2, \mu_3, \mu_4)$	$(\mu_1 * \mu_2 * \mu_3 * \mu_4)$
A	7.05	0.5	11.5	0.9	7.10	0.5	9.5	0.9	0.50	0.20
B	6.95	1.0	15.0	1.0	7.10	0.5	11.0	1.0	0.50	0.50

- to investigate whether the fuzzy system would offer any improvement in performance over the crisp system in its interpretation of results.

D. Formulation of the Initial Fuzzy Model

The crisp expert system is a forward-chaining classification system, which (after the validation phase) is based on four main input variables; arterial pH (pH_A), arterial base deficit (BD_A), venous pH (pH_V), and venous base deficit (BD_V). It was decided to restrict the initial FES only to the interpretation of *true paired* samples (samples verified as being an arterial and venous pair with validated pH_A, BD_A, pH_V, and BD_V parameters) as these rules represented a self-contained subset of the crisp system. There were 21 such crisp rules operating on the four input parameters which needed conversion into equivalent fuzzy rules. The initial choices were as follows.

1) Inference Methodology: As this application followed the conventional fuzzy control application in having crisp input variables and no internal fuzzy chaining, the Mamdani model of inference was used. All fuzzy rules were of the form "IF X_1 is A_1 THEN Y_1 is B_1" where A and B are fuzzy sets. The $\min$ operator was used for implication throughout. Although the production of complex fuzzy outputs with shape information that could be used to infer a confidence in the output was one of the original reasons for implementing a FES, it was necessary to obtain crisp outputs for the purposes of evaluation of the fuzzy model. As an arbitrary choice, center-of-gravity (centroid) defuzzification was used to produce crisp values on an arbitrary scale of 0–100 for each fuzzy output variable.

2) Linguistic Variables and Fuzzy Terms: Each of the four input parameters was assigned a linguistic variable and examination of the crisp rules showed that each could naturally be divided into three fuzzy terms corresponding to meanings of *low*, *medium*, and *high*. Two output fuzzy variables were used, severity of acidemia (*acidemia*) and duration of acidemia (*duration*). From the crisp rules it was determined that the acidemia variable had five terms in its term set: *severe*, *moderate*, *significant*, *mild*, and *none*; the duration variable had three terms: *chronic*, *intermediate*, and *acute*.

3) Rule Set: The rules for the crisp expert system were obtained as a result of knowledge elicitation sessions with several leading clinicians skilled in umbilical cord blood acid-base analysis and had been carefully refined to form a complete and consistent set of classifiers. As the crisp system had reached clinical implementation, the rule set was taken to be acceptable and was, therefore, recoded directly into fuzzy equivalents. The *fairly* hedge was taken as the square-root operator and the *very* hedge was taken as the square operator.

4) Fuzzy Operators: The probabilistic family of operators was chosen for the reason that all the fuzzy rules feature conjunction of the four linguistic variables such as

IF pH_A is low AND BD_A is $high$ AND
pH_V is low AND BD_V is $high$
THEN $acidemia$ is $severe\ metabolic.$

If the min operator is used for conjunction, the overall truth value of such a rule will obviously be determined solely by its lowest membership. For example, if the rule above is presented with the values shown in Table I, the min operator would produce the same overall truth for both cases despite the fact that case B is evidently (to an obstetric clinician) worse than case A, as the arterial pH is lower and both base deficits are higher. The probabilistic operator mimics the reasoning of the clinician in considering all the parameters so that the overall truth for B is higher.

5) Membership Functions: If the membership of any term plateaus at 1.0, then the same effect as shown in Table I can be demonstrated. The simplest way to avoid this problem is to have *sigmoid* memberships that approach 1.0 asymptotically. All fuzzy sets were modeled with *sigmoid* membership functions; left-edge sets were modeled by a decreasing *sigmoid*

$$\mu(x) = \frac{1}{1 + e^{(x-m)/\alpha\sigma}} \tag{7}$$

right-edge sets were modeled by an increasing *sigmoid*

$$\mu(x) = \frac{1}{1 + e^{(m-x)/\alpha\sigma}} \tag{8}$$

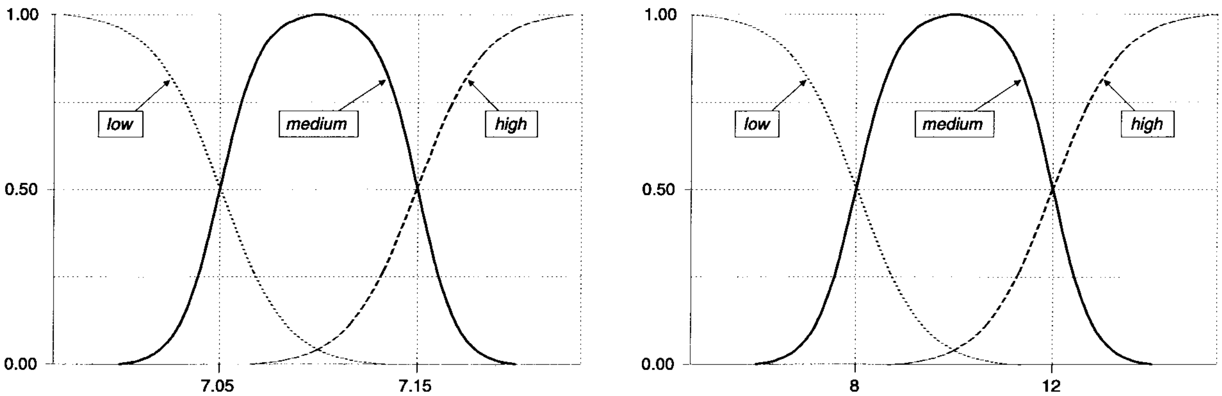
and middle sets were modeled by a combination of two *sigmoids*

$$\mu(x) = \frac{1}{(1 + e^{(x-m-\sigma/2)/\alpha\sigma})(1 + e^{(m-\sigma/2-x)/\alpha\sigma})} \tag{9}$$

where m is the location of the $\mu = 0.5$ value of the *sigmoid* for left and right sets or the location of the center value for middle sets; σ is the width parameter of the *sigmoid* corresponding to the width at $\mu = 0.5$ for middle sets and α is a constant of $\approx$0.1 governing the slope of the *sigmoids* such that the maximum value of a middle set is near to unity.

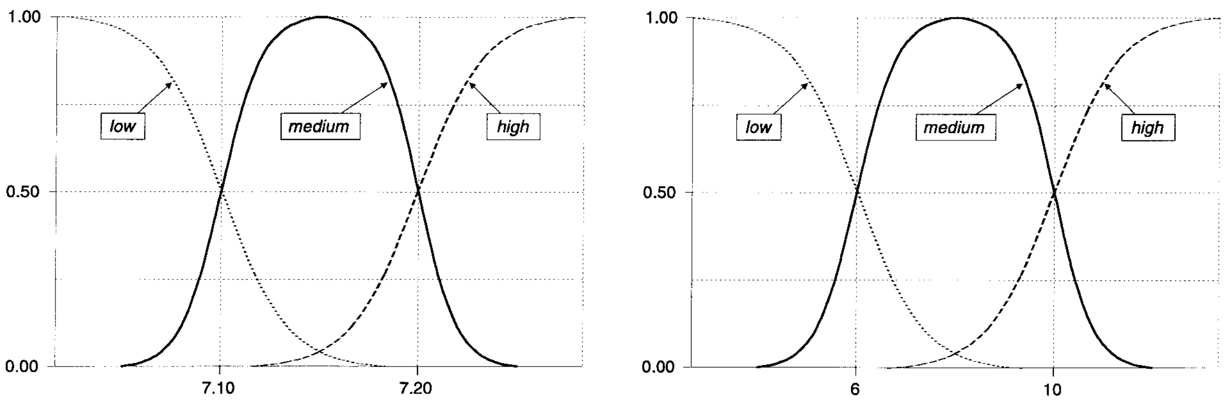
These four fuzzy input variables had the position and width of their terms determined by the cutoffs encoded into the crisp rules. Thus, for example, arterial pH fuzzy variable had its transition from *low* to *medium* at 7.05 as this value is used throughout the crisp rules. The term sets for each fuzzy input variable were generated from (7)–(9) using the universe of discourse and parameters shown in Table II and are shown in Figs. 1 and 2. The output variables were also modeled with *sigmoid* membership functions with the base variable and

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.



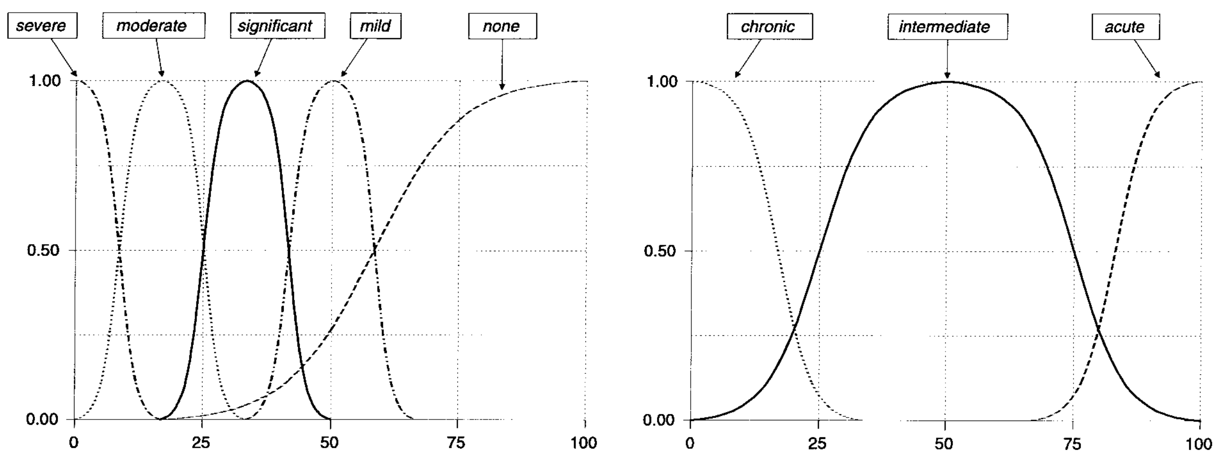
(a) (b)

Fig. 1. Term sets of (a) the arterial pH (pH_A) and (b) the arterial $\mathrm{BD_{ecf}}$ (BD_A) fuzzy input variables.



(a) (b)

Fig. 2. Term sets of (a) the venous pH (pH_V) and (b) the venous $\mathrm{BD_{ecf}}$ (BD_V) fuzzy input variables.



(a) (b)

Fig. 3. Term sets of (a) the acidemia and (b) the duration fuzzy output variable.

crossover of each term determined arbitrarily on a universe of discourse of $0 \cdots 100$. The term sets for the output variables are shown in Fig. 3.

E. Determination of the Evaluation Function

The two crisp outputs of the FES were combined into a single number to represent that duration of acidemia is of secondary importance to severity of acidemia as given by

$$\text{condition} = \text{acidemia} + \frac{\text{duration}}{10}. \tag{10}$$

The FES reanalyzed the 10 000 *true paired* samples from the database collected at Plymouth and the output was compared to the crisp system. The crisp classification had produced a category in the range 80–120. Although these category numbers were originally chosen arbitrarily, they were designed to correspond to an ordered scale in which 80 was the worst outcome (severe metabolic acidemia in both vessels) and 120 was the best outcome (normal range in both vessels). It was immediately apparent that the classification produced by the FES differed from that of the

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

TABLE II
PARAMETERS USED TO GENERATE INITIAL *SIGMOID* MEMBERSHIP FUNCTIONS FOR INPUT VARIABLES

		low		*mid*		*high*	
Variable	*Universe*	m	σ	m	σ	m	σ
pH_A	6.60 . . . 7.60	7.05	0.15	7.10	0.10	7.15	0.15
BD_A	0 . . . 20	8	6	10	4	12	6
pH_V	6.60 . . . 7.60	7.10	0.15	7.15	0.10	7.20	0.15
BD_V	0 . . . 20	6	6	8	4	10	6

TABLE III
AN EXAMPLE OF THE DIFFICULTY OF ORDERING ACID-BASE RESULTS

Results	pH_A	BD_A	pH_V	BD_V
A	6.94	11.6	6.97	11.8
B	6.87	8.5	7.11	9.4

crisp expert system, in that the ordering of results differed greatly.

An evaluation function utilizing the Spearman rank-order correlation coefficient described earlier was used to determine which order of results was the most appropriate by comparing the crisp and FES's to practising clinicians. As it was clearly impractical to get the clinicians to order all 10 000 results or even the approximately 400 abnormal results, an extract of 50 results from the 400 abnormals was chosen. The set of 50 results was selected to cover a wide range of categories by randomly selecting a number from each of the crisp categories. The clinicians were asked to rank these 50 results in order from "worst" to "best" in terms of likelihood of the infant having suffered asphyxial damage. Two clinicians involved in the creation of the rules and four clinicians considered experienced in the interpretation of cord acid-base results were recruited to take part in the comparison study. They consisted of one Professor of Physiology, one Consultant, one Senior Registrar, two Clinical Research Fellows, and a Senior Midwife.

The difficulty of this task can be illustrated by the example in Table III. In simple terms, as the pH in both vessels decreases, the infant's condition worsens and, as the BD_{ecf} in both vessels increases, the infant's condition worsens. However, the four-dimensional data must be considered in parallel so that in Table III result A might be considered worse as both the base deficits are higher than B or B might be considered worse as the arterial pH is lower than A. Additionally, the clinicians' orderings and each of the expert system's orderings were validated by comparing the results against the actual clinical outcome as ordered first by Apgar score at 5 min and then by Apgar score at 1 min.

F. Performance of the Initial Fuzzy Expert System

The results of the agreement with clinicians of the initial FES ($fuzzy^0$) is shown in Table IV. It can be seen that the average interclinician correlation $\text{ave}(\rho_s)$ is very high (0.91), indicating that the clinicians agreed with each other very well on the order of results. The crisp expert system correlated reasonably well with the clinicians (0.80), but the performance of the FES was significantly worse.

TABLE IV
RESULTS OF CLINICIANS', CRISP, INITIAL, AND TUNED FUZZY EXPERT SYSTEMS' AGREEMENT WITH CLINICIANS

Agreement	*Corr.* (ρ_s)	*Sig.* (p)
clinicians ⇔ clinicians	0.91	$\ll 0.001$
crisp system ⇔ clinicians	0.80	$\ll 0.001$
fuzzy^{0}system ⇔ clinicians	0.67	$\ll 0.001$
***fuzzy**system ⇔ clinicians**	0.93	$\ll 0.001$

TABLE V
RESULTS OF CLINICIANS', CRISP, INITIAL, AND TUNED FUZZY EXPERT SYSTEMS' AGREEMENT WITH OUTCOME (APGAR^5, APGAR^1)

Agreement	*Corr.* (ρ_s)	*Sig.* (p)
clinicians ⇔ outcome	0.47	$\ll 0.001$
crisp system ⇔ outcome	0.39	$\ll 0.001$
fuzzy^{0}system ⇔ outcome	0.17	≈ 0.001
***fuzzy**system ⇔ outcome**	0.51	$\ll 0.001$

The agreement of $fuzzy^0$ with outcome is shown in Table V. Given that other factors affect the Apgar score, the precise level of clinicians' agreement with outcome is not particularly important, but the fact that there was significant correlation beyond chance indicates that the clinicians' ordering did reflect actual clinical outcome. The important point is that the crisp expert system performed with a level close to the clinicians, but again the initial fuzzy system performed significantly worse.

IV. TUNING THE FUZZY EXPERT SYSTEM

A. Tuning Parameters

As the FES had performed badly in comparison with the clinicians and worse than the crisp system, some of the fundamental assumptions that were made in the original formulation of the fuzzy model were reconsidered. The usual method of tuning the fuzzy model would be to adjust the shape, location, and width of the membership function for each fuzzy term. Although arbitrary changes to individual membership functions are appropriate in fuzzy control type applications, the changes made to membership functions in this application were restricted in order to keep the model interpretable from a clinical knowledge level. The following changes to the fuzzy model were investigated:

- the standard "Zadeh" family of operators was used with conjunction as $\min(a, b)$, disjunction as $\max(a, b)$, and negation as $1 - a$ rather than the probabilistic family;
- linear membership functions with the same location and crossover as the sigmoid membership functions described in (7)–(9);
- the hedges *very* and *fairly* were introduced into the rules.

The reasons for the poor fuzzy performance were investigated by examination of which input variables were most

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

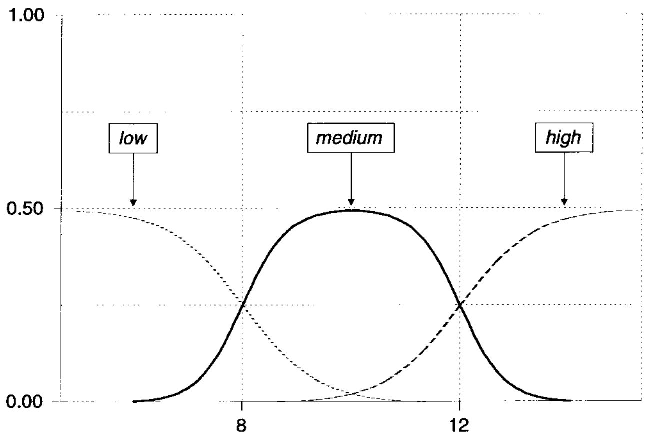


Fig. 4. Fuzzy base deficit variable with reduced membership heights.

highly correlated with the clinicians and each expert system. It was found that the clinicians and crisp expert system correlated most highly with pH_A, pH_V. In contrast, the FES correlated most highly with BD_A, BD_V. Thus, it appeared that the fuzzy system weighted the base deficit parameters more than pH in comparison to the clinicians and crisp expert system. It was decided, therefore, to attempt to improve the FES's performance by adjusting the relative weighting of the base deficit parameters in comparison to the pH parameters. A number of ways to do this were investigated:

- one new rule of the form "IF pH_A is *low* AND pH_V is *low* THEN *acidemia* is *severe*" was added as all the other rules of the FES feature a combination of pH_A, pH_V, BD_A, and BD_V, and so depend on both parameters:

* the membership functions of each of the four input variables were shifted;

* the height of pH and BD_{ecf} membership functions were decreased so that the fuzzy sets were not normalized, thus altering the influence of pH and base deficit in each rule (Fig. 4).

Initially, each of these changes was made in isolation where the biggest effect was found to be in the subnormalization of the base deficit membership functions (~0.20 performance increase) followed by the introduction of the new rule (~0.10 performance increase). The use of "Zadeh" operators and triangular membership functions both resulted in ~0.02 performance *decrease*. It was soon realized, however, that not only were these alterations not independent, but also that systematic alteration of all parameters in combination would not be computationally possible.

B. Application of the Tuning Algorithm

The effects of making these changes simultaneously were investigated by applying the SA algorithm detailed in Section II-C—those marked with a • bullet character were discrete variable changes and those marked with a $*$ bullet character were continuous variable changes. The $K = 50$ test cases were used as the target set for the FES and the minimized evaluation function was $f = 1 - \rho_s$ where ρ_s is the Spearman rank correlation of the output of the FES against the clinicians' rankings by (6) in Section II-D. Thus, for this application

$$\begin{aligned}\boldsymbol{x} = (&x_1 : operator\ family, \cdots\\ &x_2 : membership\ function\ shape, \cdots\\ &x_3 : rule\ set, \cdots\\ &x_4 : \text{very}\ hedging, x_5 : \text{fairly}\ hedging)\\ \boldsymbol{y} = (&y_1 : pH_A\ offset, y_2 : BD_A\ offset, \cdots\\ &y_3 : pH_V\ offset, y_4 : BD_V\ offset, \cdots\\ &y_5 : pH_A\ height, y_6 : BD_A\ height, \cdots\\ &y_7 : pH_V\ height, y_8 : BD_V\ height).\end{aligned}$$

The variables were encoded as follows:

$$\begin{aligned}x_1 &= \{0 \equiv \text{probabilistic operators},\\ &\quad 1 \equiv \text{min-max operators}\}\\ x_2 &= \{0 \equiv \text{sigmoid membership functions},\\ &\quad 1 \equiv \text{linear membership functions}\}\\ x_3 &= \{0 \equiv \text{original rule set},\\ &\quad 1 \equiv \text{extra rule included in rule set}\}\\ x_4 &= \{0 \equiv \text{no } very \text{ hedges}, \cdots\\ &\quad \sum_{i=1}^{21} 2^i \equiv very \text{ hedge for acidemia in Rule } i\}\\ x_5 &= \{0 \equiv \text{no } fairly \text{ hedges}, \cdots\\ &\quad \sum_{i=1}^{21} 2^i \equiv fairly \text{ hedge for acidemia in Rule } i\}\end{aligned}$$

$$\begin{aligned}y_1 &= [6.50, 7.50]\\ y_2 &= [0, 20]\\ y_3 &= [6.50, 7.50]\\ y_4 &= [0, 20]\\ y_5 &= [0, 1]\\ y_6 &= [0, 1]\\ y_7 &= [0, 1]\\ y_8 &= [0, 1]\end{aligned}$$

The fuzzy model was then tuned automatically with Algorithm C by evaluating the cost function as follows:

1) Generate fuzzy model input term memberships as shown in Table VI.
2) If $x_2 = 0$, then use *sigmoid* membership functions or else use linear membership functions.
3) Reduce height of three pH_A terms by y_5, reduce height of three BD_A terms by y_6, reduce height of three pH_V, terms by y_7, reduce height of three BD_V terms by y_8.
4) Run the fuzzy model on K input data:
   a) if $x_3 = 0$, then use original rule set or else use the rule set with additional rule;
   b) if $x_4 > 0$, then insert a *very* hedge operator into specified rules;
   c) if $x_5 > 0$, then insert a *fairly* hedge operator into specified rules;

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

TABLE VI
PARAMETERS USED TO GENERATE TUNABLE *SIGMOID* MEMBERSHIP FUNCTIONS FOR INPUT VARIABLES

	low		*mid*		*high*	
Variable	m	σ	m	σ	m	σ
pH_A	y_1	0.15	$y_1 + 0.05$	0.10	$y_1 + 0.10$	0.15
BD_A	$y_2 - 4$	6	$y_2 - 2$	4	y_2	6
pH_V	y_3	0.15	$y_3 + 0.05$	0.10	$y_3 + 0.10$	0.15
BD_V	$y_4 - 4$	6	$y_4 - 2$	4	y_4	6

TABLE VII
AN EXAMPLE OF SIX SEPARATE RANKINGS, WITH THE ROW-WISE SUM AND *OPTIMAL* RANKING

	Clinician							
Case	1	2	3	4	5	6	$\sum_1^6$	r_{opt}
A	1	1	2	2	1	2	9	1
B	2	3	1	3	2	3	14	3
C	3	2	3	1	3	1	13	2

d) if $x_1 = 0$, then use probabilistic operators or else use $min\text{-}max$ operators;
e) defuzzify output variables and calculate single crisp output by (10).

5) Rank K outputs from $1 \cdots K$ in order from worst to best.
6) Calculate ρ_s for each clinician by (6), and then calculate the average FES agreement $\overline{\rho_s}$.
7) Cost function $= 1 - \overline{\rho_s}$.

In this case, the discrete variables $x_1 \cdots x_5$ effectively represent alternative structures for the fuzzy model and the continuous variables $y_1 \cdots y_8$ represent adjustable parameters within the model. For example, if the x_3 variable was randomly changed by the algorithm from zero to one, then the additional rule operated within the model. If x_3 reverted to zero then the additional rule was not operated.

An approximation to the optimal cost function f_{opt} was obtained by calculating the sum of the ranks of the six clinicians for each of the K cases and then reranking these (row-wise) sums. For example, if six clinicians each ranked three cases A, B, and C as shown in Table VII, the optimal ranking would be $A = 1, B = 3$, and $C = 2$, as shown in the right-hand column. This *ranking of the summed ranks* can be thought of as the best *consensus* view that gave a value of $\rho_s(\text{opt}) \approx 0.97$ and, therefore, $f_{\text{opt}} = 1 - \rho_s(\text{opt}) \approx 0.03$. However, it is not necessarily the case that *any* configuration of the fuzzy model could attain f_{opt}, so in practice the average interclinician agreement $\text{ave}(\rho_s) = 0.91$ was taken as $\rho_s(\text{min})$. Thus, $f < 1 - \rho_s(\text{min}) \approx 0.09$ was taken as an acceptible performance.

The initial temperature, number of repeats and λ parameter (step 2, Algorithm B) were found by trial and error. The tunable parameters were initialized in a random configuration and the annealing algorithm was run using a cooling schedule whereby the temperature was reduced by a factor of $\omega = 0.95$ until it reached $\frac{1}{100}$th its initial value (90 iterations). The number of repeats was typically around 300 per iteration, so that a total of ≈27 000 annealings were performed. Each evaluation of the cost function by running the fuzzy model took ∼1 s on a 200 MHz Pentium PC and, therefore, each annealing trial took ≈8 h. At the completion of each run, the final configuration of the fuzzy model and its performance in terms of the lowest cost function was stored. Multiple runs were undertaken from different random starting points in case the annealing had located a local minima.

C. Performance of the Tuned Fuzzy Expert System

The biggest effect on the performance of the FES was found to be the following.

1) The subnormalization of the base deficit term sets: the subnormalization was carried out on each base deficit variable independently, affecting all the terms equally. Very slight differences in the performance could be obtained through having the height of the two variables slightly unequal, but overall the best effect was achieved by reducing the height of both variables' term sets to 0.4.
2) The introduction of the extra rule specified above: a very large number of other rule changes could have been considered such as creating new rules and/or deleting the existing rules, but this was not attempted in order to keep the model clinically interpretable.

The choice of probabilistic operators was found to have a small beneficial effect over the $\max/\min$ operators, as was the use of *sigmoid* membership functions over *triangular* functions. The introduction of hedges was found to have a negligible effect. Thus, the tuned FES has membership functions identical in shape and location to those shown in Figs. 1 and 2, but with the maximum height of all the base deficit terms reduced to 0.4. No alteration of the output variable membership functions was attempted. After introduction of the single additional rule and reduction of the base deficit membership heights to 0.4, the final results shown as $fuzzy^*$ in Tables IV and V were obtained.

It can be seen that the tuned FES $(fuzzy^*)$ achieved an excellent agreement with the clinicians and matched the clinicians in its agreement with outcome. Its modified performance was better than the crisp expert system and effectively indistinguishable from the clinicians. A graph of the six clinicians rankings plotted against the $fuzzy^*$ rankings is shown in Fig. 5. If perfect agreement had been found, a straight line through the origin would have been obtained. For example, for the case that the $fuzzy^*$ system ranked the worst (ranking 1), all the clinicians also ranked the worst (ranking 1), and all the points are superimposed at (1, 1). On the other hand, for the case that the $fuzzy^*$ system ranked 25th worst, the clinicians ranked as 14th, 16th, 17th, 18th, 19th, and 20th—these points therefore appear at (25, 14), (25, 16), (25, 17), (25, 18), (25, 19), and (25, 20).

An external criterion known as the *Apgar score* was available for subsequent post-tuning validation of the final FES. There are many other clinical factors apart from umbilical acid-base results that affect the Apgar score, so that high correlation between the two would not be expected clinically. Despite this, *better* interpretation of acid-base results would

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

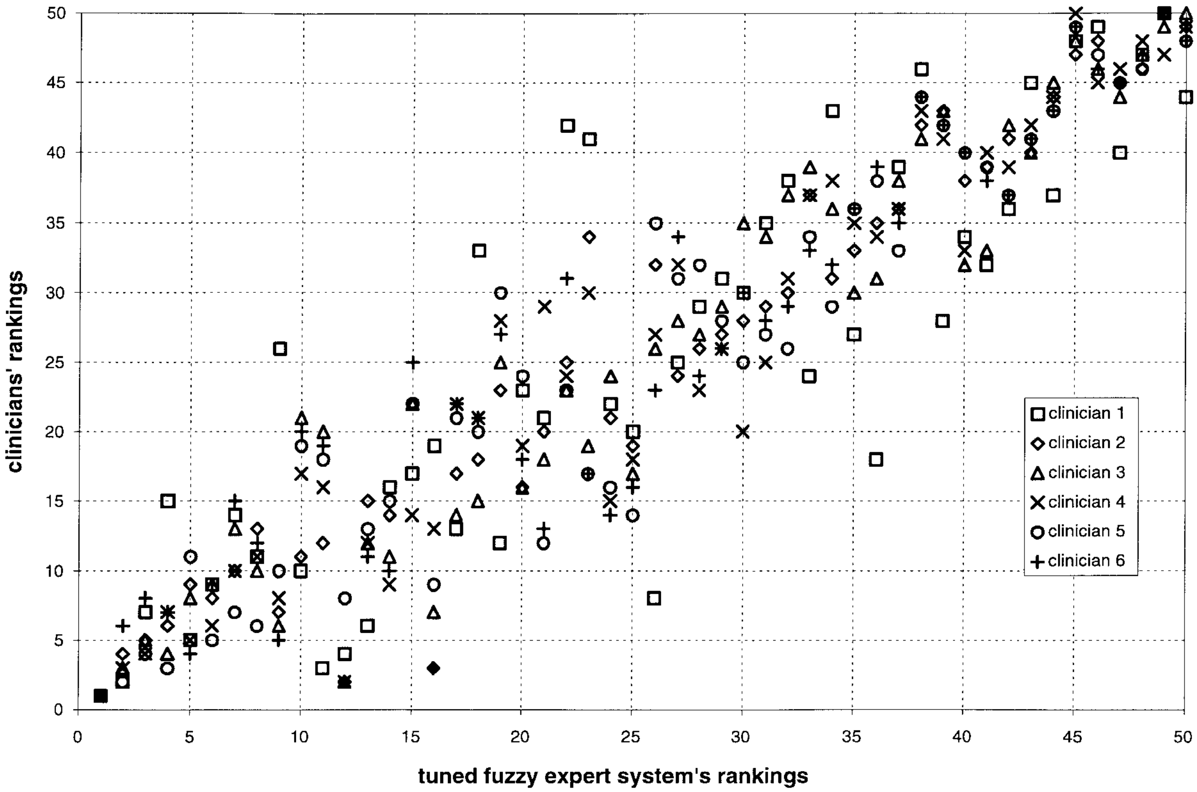


Fig. 5. Graph of six clinicians' rankings against the optimized FES.

TABLE VIII
VALIDATION OF EXPERT SYSTEMS WITH OUTCOME (APGAR^5, APGAR^1) FOR ALL CASES ($n = 9\,003$)

Agreement	*Corr.* (ρ_s)	*Sig.* (p)
crisp system ⇔ outcome	0.21	$\ll 0.001$
fuzzy^0 system ⇔ outcome	0.14	$\ll 0.001$
fuzzy* **system ⇔ outcome**	0.24	$\ll 0.001$

TABLE IX
VALIDATION OF EXPERT SYSTEMS WITH OUTCOME (APGAR^5, APGAR^1) FOR ALL ABNORMAL CASES ($n = 383$)

Agreement	*Corr.* (ρ_s)	*Sig.* (p)
crisp system ⇔ outcome	0.44	$\ll 0.001$
fuzzy^0 system ⇔ outcome	0.26	< 0.001
fuzzy* **system ⇔ outcome**	0.52	$\ll 0.001$

be expected to have *better* correlation with Apgar score. Thus, the *relative* correlation of the expert systems against the entire database of cases with *true paired* samples where the Apgar scores had been recorded ($n = 9\,003$) was used to evaluate their performance. The results in Table VIII showed that the $fuzzy^*$ indeed performed the best of the three systems, albeit with fairly low overall correlation. This lower correlation is due to the fact that the vast majority of cases (~98%) have normal acid-base results and normal Apgar scores. Thus, the agreement was also calculated against the Apgar scores for all cases with abnormal acid-base status ($n = 383$, defined by the crisp cutoffs $pH_A < 7.05$, $BD_A \geq 12$ mmol.l^{-1}, $pH_V < 7.10$, and $BD_V \geq 10$ mmol.l^{-1}) and the results in Table IX were obtained. For these novel abnormal cases, the $fuzzy^*$ system again achieved the highest correlation of all three systems and at a level similar to that found during tuning (Table V).

V. DISCUSSION AND CONCLUSIONS

The initial performance of the FES was poor—worse than the original crisp system—not only when compared to clinicians in the test that was designed, but also when compared to the most popular form of current overall neonatal outcome assessment (the Apgar score). The reason for this poor performance is not clear. In the crisp system the rule set is nonoverlapping and exclusive; each set of input parameters will fire a single rule. It appeared that the multiple interaction of the fuzzy rules (especially in the intermediate cases) caused its performance to degrade, due to weighting the base deficit parameters too highly (see Section IV). Ultimately, the reason for the poor performance could only be investigated through lengthy discussions with the clinicians on a case by case basis to identify situtations where the crisp system out-performed the fuzzy system, and then to examine the internal processing to establish "incorrect" behavior. Unfortunately, lack of availability of the clinicians prevented such an investigation.

An encouraging aspect of this study was that the clinicians did form a good body of agreement between themselves—not always the case in studies involving *expert* clinicians. The performance of the optimized FES was excellent—effectively indistinguishable from the clinicians and the best of the expert systems in the final validation test against Apgar scores.

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

Obviously, the FES has been tailored to perform well on the training set of 50 cases and would have to be retested against clinician's opinion on novel cases to ensure that the results can be generalized. However, the advantage of the approach described over a neural network approach is that the final fuzzy system can be analyzed and discussed with the (expert) clinicians to identify whether its rule set is acceptable on the meta level. The two changes made to produce the final fuzzy model can both be justified from clinical knowledge and the resultant fuzzy model is attractive through its ability to vary the relative importance of the pH and base deficit parameters simply through alteration of the height of the base-deficit term sets.

The use of subnormal membership functions to adjust the relative weighting of fuzzy terms is a novel and promising general purpose technique. The decision to change the height of membership functions was made in an attempt to rectify the apparent over emphasis that the FES placed on the base-deficit parameters. In the vast majority of applications and in most theoretical work on fuzzy sets, it is assumed that membership functions should be normalized. While it is true that most (if not all) clinicians would label a base deficit of, for example, 25 as *high* with a truth of 1.0; it is questionable whether the truth that the statement conveys is *as true* as the statement that a pH of 6.60 is *low*—especially when the statements are considered in the context of possible damage having occurred to a neonate. In other words, the context in which a particular linguistic term is used and its relationship to other linguistic variables in the same context can alter its maximum attainable truth value.

If one accepts the central tenet of fuzzy logic—that truth can have different linguistic values with shape and form to represent different human concepts of truth—then we see no problem with accepting that certain statements will never be *as true* as others within a specified domain. The use of subnormal fuzzy sets in a linguistic variable allows the truth of statements concerning that variable to be weighted within a domain in a natural and consistent manner. Any internal change to a fuzzy model that alters the input–output mapping can be said to change the meaning of the rules embodied in the model. There is almost certainly an alternative formulation of the fuzzy model that would achieve as good a performance as the system presented here, but without needing to use subnormal sets. Such a model may feature a set of rules in which the pH parameters and $\mathrm{BD_{ecf}}$ parameters are separated out with rules concerning base deficits having an overall weighting factor, but this representation may not be as appealing to the clinician.

Indeed there are many other changes to the initial fuzzy system that could have been made in addition to or instead of subnormal membership functions. Examples of such changes include alternative or additional rule sets, different numbers of input or output membership functions, alteration of the width and location of individual membership functions, or alternative defuzzification strategies. Such changes were not tried, as they might not result in a model that was interpretable by the clinicians, whereas the final model produced does represent an internally consistent model of knowledge for the human clinicians.

The SA algorithm is a simple general-purpose technique that can optimize a fuzzy model used in an FES. Other optimization techniques could have been used and would have been equally valid. SA is not necessarily the quickest or most efficient form of global optimization in terms of function evaluations, but it is easy to apply to both discrete and continuous variables and, thus, can simultaneously tune the structure and parameters of a model. If pragmatic considerations of speed (in terms of coding effort) and ease of getting a fuzzy system working to a high-performance level outweigh the theoretical drawbacks of the SA approach, then the method is extremely useful.

Appendix A
Discrete Simulated Annealing Algorithm A

Let $\boldsymbol{x}$ be a vector of m discrete variables $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)$, where each $x_i \in X_i$, the set of all possible values of the ith discrete variable. Let $f(\boldsymbol{x})$ be the cost function being optimized (minimized).

1) Let $\boldsymbol{x}_o$ be an arbitrary initial (either random or specified) starting point, let T be the initial temperature, and let L_d be the initial number of repetitions.
2) Repeat L_d times:
   a) generate a trial point $\boldsymbol{x}^*$ from the set of all possible solutions;
   b) if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}_o)$, set $\boldsymbol{x}_o = \boldsymbol{x}^*$;
   c) if $f(\boldsymbol{x}^*) > f(\boldsymbol{x}_o)$, then if $\exp(f(\boldsymbol{x}_o) - f(\boldsymbol{x}^*)/T) > \mathrm{random}[0, 1)$, set $\boldsymbol{x}_o = \boldsymbol{x}^*$.
3) If *stopping criterion*, stop.
4) Decrease temperature T and adjust repetitions L_d according to annealing schedule; go to step 2.

Appendix B
Continuous Simulated Annealing Simplex Algorithm B

Let $\boldsymbol{y}$ be a vector of n continuous variables $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$, where each $y_i \in \Re$ with bounds $[y_i', y_i'']$. Let $f(\boldsymbol{y})$ be the cost function being optimized (minimized), let $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n, \boldsymbol{y}_{n+1}$ be $n + 1$ vectors representing a simplex in the n-dimensional space. Let $F_j = f(\boldsymbol{y}_j), G_j = f(\boldsymbol{y}_j) - \boldsymbol{T}\ \ln(\mathrm{random}(\boldsymbol{0}, \boldsymbol{1}))$ (F_j increased by a thermal fluctuation factor), and $H_j = f(\boldsymbol{y}_j) + \boldsymbol{T}\ \ln(\mathrm{random}(0, 1))$ (F_j decreased by a thermal fluctuation factor). If any $\boldsymbol{y}$ is generated with y_i outside bound y_i' or y_i'', use (3) to evaluate the function.

1) Let $\boldsymbol{y}_o$ be an arbitrary initial (either random or specified) starting point, let T be the initial temperature and let L_c be the initial number of repetitions.
2) Generate the initial simplex: let $\boldsymbol{y}_i = \boldsymbol{y}_o + \lambda \boldsymbol{e}_i$ for $i = 1 \cdots n$ where the $\boldsymbol{e}_i$'s are n unit vectors and λ is a constant corresponding to the characteristic length of the problem and let $\boldsymbol{y}_{n+1} = \boldsymbol{y}_o$.
3) Evaluate the $n + 1$ cost functions, $F_j = f(\boldsymbol{y}_j)$ for $j = 1 \cdots n + 1$.
4) Repeat L_c times:
   a) sort simplex points in order from best to worse (with a thermal fluctuation) such that $G_1 \leq G_2 \leq \ldots \leq G_n \leq G_{n+1}$;

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

b) calculate the centroid of n best points in $n+1$-dimensional space

$$\boldsymbol{c} = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{y}_j$$

c) Reflect the worst point through the "facing" centroid point $\boldsymbol{y}_r = \boldsymbol{c} + (\boldsymbol{c} - \boldsymbol{y}_{n+1})$;

d) if $H_r < F_1$ expand the reflected new best point $\boldsymbol{y}_e = \boldsymbol{c} + 2(\boldsymbol{c} - \boldsymbol{y}_{n+1})$; then if $H_e < H_r$, set $\boldsymbol{y}_{n+1} = \boldsymbol{y}_e$, or else set $\boldsymbol{y}_{n+1} = \boldsymbol{y}_r$;

e) if $F_1 \leq H_r \leq F_n$ set $\boldsymbol{y}_{n+1} = \boldsymbol{y}_r$;

f) if $F_n < H_r < F_{n+1}$ contract the reflected simplex $\boldsymbol{y}_c = \boldsymbol{c} + \frac{1}{2}(\boldsymbol{y}_r - \boldsymbol{c})$;

g) if $H_r \geq F_{n+1}$ contract the original simplex $\boldsymbol{y}_c = \boldsymbol{c} + \frac{1}{2}(\boldsymbol{y}_{n+1} - \boldsymbol{c})$;

h) if (and only if) a contraction was attempted, if $H_c < \min(H_r, F_{n+1})$, then the contraction succeeded; set $\boldsymbol{y}_{n+1} = \boldsymbol{y}_c$ or else contract all points of the simplex toward the best point, set $\boldsymbol{y}_j = \boldsymbol{y}_1 + \frac{1}{2}(\boldsymbol{y}_j - \boldsymbol{y}_1)$ for $j = 2 \cdots n+1$.

5) If *stopping criterion*, stop.
6) Decrease temperature T and adjust repetitions L_c according to annealing schedule: go to step 4.

APPENDIX C
SIMULATED ANNEALING FUZZY TUNING ALGORITHM C

Let $\boldsymbol{x}$ be a vector of m discrete variables $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)$ and let $\boldsymbol{y}$ be a vector of n continuous variables $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$. Let $f(\boldsymbol{x}, \boldsymbol{y})$ represent the cost function being minimized obtained by running the fuzzy model parameterized by $\boldsymbol{x}$ and $\boldsymbol{y}$ on K sets of input data and evaluating the K sets of output data against required output (for example, through the use of an agreement measure as described in Section II-D).

1) Collect K samples of training data with inputs and desired output.
2) Select the m discrete variables, $x_1, \cdots, x_m$ used to construct the fuzzy model.
3) Select the n continuous parameters $y_1, \cdots, y_n$ used to adjust the fuzzy model.
4) Initialize the fuzzy model: let $\boldsymbol{x}_o$ be the initial starting configuration of discrete variables, let $\boldsymbol{y}_o$ be the initial starting configuration of continuous parameters, let T be the initial annealing temperature, let L_d be the initial number of discrete repetitions, and let L_c be the initial number of continuous (simplex) repetitions.
5) Initialize the simplex for continuous optimization as in steps 2–3 of Algorithm B.
6) Decide whether to make discrete or continuous adjustment to the fuzzy model.
7) If discrete adjustment, perform L_d discrete annealing repetitions to tune $\boldsymbol{x}$ in $f(\boldsymbol{x}, \boldsymbol{y})$ as in step 2 of Algorithm A.
8) If continuous adjustment, perform L_c simplex annealing repetitions to tune $\boldsymbol{y}$ in $f(\boldsymbol{x}, \boldsymbol{y})$ as in step 4 of Algorithm B.
9) If *stopping criterion*, stop.
10) Decrease temperature T and adjust repetitions L_d and L_c according to annealing schedule, go to step 6.

ACKNOWLEDGMENT

The authors would like to thank the referees for the many helpful comments made during preparation of this paper. The authors would like to thank K. Greene for his clinical support. They would also like to thank the experts who took part in the comparison study: J. Westgate, K. Rosén, M. Davies, S. Beckley, G. Hughes, and M. Harris. Thanks also go to N. Hughes for inspiration and helpful guidance on the use of the SA algorithm.

REFERENCES

[1] C. Hughes, "The representation of uncertainty in medical expert systems," *Medical Informatics*, vol. 14, pp. 269–279, 1989.

[2] L. A. Zadeh, "The role of fuzzy logic in the management of uncertainty in expert systems," *Fuzzy Sets Syst.*, vol. 11, pp. 199–227, 1983.

[3] K. P. Adlassnig, "Fuzzy set theory in medical diagnosis," *IEEE Trans. Syst., Man, Cybern.*, vol. 16, no. 2, pp. 260–265, Mar. 1986.

[4] M. E. Cohen and D. L. Hudson, "The use of fuzzy variables in medical decision making," in *Fuzzy Computing*, M. M. Gupta and T. Yamakawa, Eds. North Holland, The Netherlands: Elsevier, 1988, pp. 263–271.

[5] M. Halim, K. M. Ho, and A. Liu, "Fuzzy logic for medical expert systems," *Ann. Acad. Med.*, vol. 19, no. 5, pp. 672–683, 1990.

[6] K. P. Adlassnig and G. Kolarz, "CADIAC-2: Computer-assisted medical diagnosis using fuzzy subsets," in *Approximate Reasoning in Decision Analysis*, M. M. Gupta and E. Sanchez, Eds. North-Holland, The Netherlands: Elsevier, 1982, pp. 219–247.

[7] H. Watanabe, W. J. Yakowenko, Y. M. Kim, J. Anbe, and T. Tobi, "Application of a fuzzy discrimination analysis for diagnosis of valvular heart disease," *IEEE Trans. Fuzzy Syst.*, vol. 2, pp. 267–276, Nov. 1994.

[8] C. T. Lin and C. S. G Lee, "Neural-network based fuzzy logic control and decision making," *IEEE Trans. Comput.*, vol. 40, pp. 1320–1336, Dec. 1991.

[9] K. M. Lee, D. H. Kwak, and H. Lee-Kwang, "Fuzzy inference neural network for fuzzy model tuning," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, no. 4B, pp. 637–645, Aug. 1996.

[10] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 698–713, Sept. 1992.

[11] J. S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, pp. 665–685, May 1993.

[12] J. Valente de Oliveira, "A design methodology for fuzzy system interfaces," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 404–414, Nov. 1995.

[13] W. Pedrycz and J. Valente de Oliveira, "Optimization of fuzzy models," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, no. 4B, pp. 627–636, Aug. 1996.

[14] R. Rovatti and Guerrieri, "Fuzzy sets of rules for system identification," *IEEE Trans. Fuzzy Syst.*, vol. 4, pp. 89–102, May 1996.

[15] W. Pedrycz and J. Valente de Oliveira, "An algorithmic framework for development and optimization of fuzzy models," *Fuzzy Sets Syst.*, vol. 80, pp. 37–55, 1996.

[16] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[17] L. I. Kuncheva, "Pattern-recognition with a model of fuzzy neuron using degree of consensus," *Fuzzy Sets Syst.*, vol. 66, no. 2, pp. 241–250, 1994.

[18] C. F. Ball and D. A. Mlynski, "Fuzzy bi- and multi-partitioning for circuits represented by hypergraphs," *J. Circuits Syst. Comput.*, vol. 6, pp. 503–526, 1996.

[19] C. H. Sing and B. Postlethwaite, "Fuzzy relational model-based control applying stochastic and iterative methods for model identification," *Chem. Eng. Res. Design*, vol. 74, pp. 70–76, 1996.

[20] W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[21] J. M. Garibaldi, J. A. Westgate, E. C. Ifeachor, and K. R. Greene, "The development and implementation of an expert system for the analysis of umbilical cord blood," *Artif. Intell. Med.*, vol. 10, pp. 129–144, 1997.

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.

[22] F. Mandl, *Statistical Physics*, 2nd ed. New York: Wiley, 1988.
[23] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. New York: Wiley, 1989.
[24] D. Vanderbilt and S. G. Louie, "A Monte Carlo simulated annealing approach to optimization over continuous variables," *J. Computat. Phys.*, vol. 56, pp. 259–271, 1984.
[25] I. O. Bohachevsky, M. E. Johnson, and M. L. Stein, "Generalized simulated annealing for function optimization," *Technometrics*, vol. 28, no. 3, pp. 209–217, 1986.
[26] A. Corana, M. Maresi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the 'simulated annealing' algorithm," *ACM Trans. Math. Software*, vol. 13, no. 3, pp. 262–280, 1987.
[27] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.
[28] R. R. Yager, "On a general class of fuzzy connectives," *Fuzzy Sets Syst.*, vol. 4, pp. 235–242, 1980.
[29] S. Siegel and N. J. Castellan, *Nonparametric Statistics for the Behavioral Sciences*, 2nd ed. New York: McGraw-Hill, 1988.
[30] S. Arulkumaran and I. Ingermarsson, "Appropriate technology in intrapartum surveillance," in *Progress in Obstetrics and Gynecology*, J. Studd, Ed. London, U.K.: Churchill Livingstone, 1990, vol. 7, pp. 127–140.
[31] E. H. Shortliffe, "The adolescence of AI in medicine: Will the field come of age in the '90s?" *Artif. Intell. Med.*, vol. 5, pp. 93–106, 1993.
[32] O. Siggaard-Andersen, *The Acid-Base Status of the Blood*, 4th ed. Copenhagen, Denmark: Munksgaard, 1976.
[33] K. G. Rosén and K. W. Murphy, "How to assess fetal metabolic acidosis from cord samples," *J. Perinatal Med.*, vol. 19, pp. 221–226, 1991.
[34] J. A. Westgate, J. M. Garibaldi, and K. R. Greene, "Umbilical cord blood gas analysis at delivery: A time for quality data," *British J. Obstetrics Gynecology*, vol. 101, pp. 1054–1063, 1994.

Jonathan M. Garibaldi received the B.Sc. (honors) degree in physics from Bristol University, Bristol, U.K., and the M.Sc. (intelligent systems) and Ph.D. degrees from the University of Plymouth, Devon, U.K., in 1984, 1991, and 1997, respectively.

He is a Research Fellow in Intelligent Medical Systems within the School of Electronic Communication and Electrical Engineering, University of Plymouth, Devon, U.K., Having been appointed in 1997. In 1984 he spent six years in industry as a Software Engineer after which he took a leave to pursue his master's degree. After a brief return to the software industry, he re-entered the University of Plymouth as a Research Assistant in 1992. His main research interests are in intelligent systems techniques including expert systems, fuzzy theory, and neural networks and genetic algorithms, with particular emphasis on practical medical applications. He also has interests in data mining, programming languages, probability, statistics, and computability.

Emmanuel C. Ifeachor received the B.Sc. (honors) degree in communication engineering from University of Plymouth (formerly Plymouth Polytechnic), Devon, U.K., in 1980, the M.Sc. degree in communication engineering, the Diploma of Membership of Imperial College (DIC) from Imperial College, London, U.K., in 1981, and the Ph.D. degree in medical electronics from University of Plymouth in 1985.

He is the Wandel and Goltermann Professor of Intelligent Electronic Systems, Head of School of Electronic Communication and Electrical Engineering, and Head of Signal Processing and Intelligent Systems Research Group at University of Plymouth. He did his industrial training with GEC Telecommunication Ltd., Coventry, U.K., in 1978 and 1979. He served as a Research Assistant in the Department of Communication Engineering at the University of Plymouth and became a Lecturer in 1985. He has published over 80 technical papers, coauthored *Digitial Signal Processing—A Practical Approach* (Reading, MA: Addison-Wesley, 1993), and edited three books. His main research interests are signal processing and intelligent systems with applications in biomedicine, audio engineering, and telecommunications.

Dr. Ifeachor recently received the Institute of Electrical Engineers Dr. V. K. Zworykin Premium for his work on fetal electrocardiogram analysis and currently serves on the Institute's Professional Group on Biomedical Engineering.

Authorized licensed use limited to: UNIVERSITY OF NOTTINGHAM. Downloaded on February 5, 2010 at 09:20 from IEEE Xplore. Restrictions apply.