Fuzzy Methods and Medical Diagnosis

P.R. Innocent, R.I. John and J. M. Garibaldi
The Centre for Computational Intelligence
Department of Computer Science
De Montfort University, Leicester, UK
pri@dmu.ac.uk

January 29, 2004

Abstract

This paper argues that fuzzy representations are appropriate in applications where there are major sources of imprecision and/or uncertainty. Case studies of fuzzy approaches to specific problems of medical diagnosis and classification are described in support of this argument. The case studies are in the areas of categorical consistency, diagnostic monitoring and scoring. The solutions use a variety of fuzzy methods including clustering, fuzzy set aggregation and type-2 fuzzy set modelling of linguistic approximations. It is concluded that the fuzzy approach to the development of artificial intelligence in application systems is beneficial in these contexts because of the need to focus on uncertainty as a main issue.

Keywords: Diagnosis, Clustering, Monitoring, Linguistic Variables, Fuzzy Logic, Type-2 Fuzzy Sets.

1 Introduction

The development of artificial intelligence methodology has been recognised as an important requirement in complex problem solving situations. Medical diagnosis is a particularly good example because of the complexity of the human mind and body and our limited and vague knowledge of how these function. This knowledge also varies with the expertise of the user. In a systematic approach to the acquisition of domain knowledge, the analysis of human physiology/psychology quickly produces large numbers of cause-effect relations at many interacting levels of both description and function. Necessarily, the relations are poor approximations of complex dynamic systems and some account has to be made for uncertainty at this level of description. Furthermore, the information available for searching this domain knowledge for a specific diagnosis is also usually vague (at least initially) in that evidence is indirect (reported symptoms or lack of them) and observations are incomplete and inaccurate due to the stochastic nature of mind/body psychology/physiology. The level of expertise of the clinician cannot be discounted in this process. Given that speed is also important in the diagnostic procedure, we should develop techniques in artificial intelligence that can support fast, reliable and accurate diagnosis with limited and vague information.

It is important that experts are repeatedly able to classify the same situation in the same category. In this paper, we refer to the term 'categorical consistency' (defined in [28] with respect to compatibility of between data and models) as a measure of how well experts can achieve repeatable classification performance. Categorical consistency is a pre-requisite for the consultation systems such as those expert systems based on traditional knowledge acquisition techniques. However, it is important to understand that the establishment of consistent categories for classification is a result of intelligent processing by human experts and there may always be disagreement between experts especially in uncertain information conditions. Therefore, in this paper, we first consider how categories are established in a case study concerned with radiographic image classification. We will show how fuzzy representations are useful for taking uncertainty into account and can be applied to model the acquisition of knowledge by experts. Steimann and Adlassnig [26] have presented a strong case for using fuzzy sets to support heuristic methods of diagnosis based on placing a major emphasis on the uncertainty in the information in the process. This paper considers two further case studies in different diagnostic domains defined by Steimann and Adlassnig: diagnostic monitoring and scoring. We show how fuzzy representation and processing supports these activities by allowing uncertainty and imprecision to be taken into consideration.

We first informally introduce the basic concepts of type-1 and type-2 fuzzy sets. The purpose of this section is to briefly outline the basic ideas of fuzzy sets for dealing with uncertainty and imprecision in an informal manner which is sufficient only for the needs of this paper. Readers with knowledge of this may omit this section. For those who wish a more formal discussion of the basic concepts see, for example,[15] and [20].

2 Introduction to Fuzzy Sets

2.1 Type-1 Fuzzy Sets

Informally, a crisp (non-fuzzy) set is a collection of objects. Each object is either a member of a particular crisp set or not. In a fuzzy set, an object is a member of a particular fuzzy set to some degree, called a membership value. In a type-1 fuzzy set, membership values are in the range zero to one [0,1]. For example, a glass of water may be drinkable to degree 0.3. This implies that the water may not be very pleasant or have bad side effects to this degree. It is important to note that the membership degree is not the same as probability. If the probability of a glass of water is undrinkable is 0.3, the implication is that, out of 10 glasses, 3 of them are completely undrinkable and the other 7 are drinkable. That is, the classes 'drinkable' and 'undrinkable' are represented as crisp sets and an object (glass of water) exclusively belongs to one of them. An implication could be made, for example, that 3 of the glasses were filled by water from one source and the other 7 from another. In the crisp set, if there is only one glass, it will contain water that is drinkable or undrinkable. In the fuzzy case, there may be only one glass that has been judged by a tester to be perceived as being undrinkable to some degree. The rules of probability (e.g. that all relevant probabilities sum to 1) do not apply to the fuzzy sets because there are separate mathematical operators and rules which govern the union, intersection and other operations on fuzzy sets.

A more formal definition of a fuzzy set (a type-1 fuzzy set) is:

Definition 1 *For any fuzzy set A, the function μ_A represents the membership function for which $\mu_A(x)$ indicates the degree of membership that x, of the universal set X, belongs to set A and is, usually, expressed as a number between 0 and 1:*

$$\mu_A(x) : X \to [0,1].$$

For a discrete fuzzy set A, with members $x_1, \ldots x_N$ the usual notation is to write $A = \mu_1/x_1 + \mu_2/x_2 + \cdots + \mu_N/x_N$. In this case, the $+$ means union.

The determination of membership functions is difficult and people may characterise fuzzy sets by using a language term such as one from the set high, low, medium. Thus, for a particular judge, a glass of water may be drinkable to high degree, where ‘high’ is defined as a fuzzy set over the domain [0,1] and at the same time undrinkable to a ‘very low’ degree where ‘very’ is a modifier of the fuzzy set ‘low’. To arrive at an outcome for deciding whether a particular sample of water is drinkable from a large number of judges, we must use fuzzy set composition rules (for example, the simple additive model in [20]) to combine these fuzzy sets and then defuzzify the combined set into a crisp value to determine the overall assessment of the drinkability of a particular sample. Now suppose that we invent a device that measures an estimate of ‘water goodness’. In a crisp approach, we would have a simple threshold for this value above which we would assign the sample to the ‘undrinkable’ set and below which we would assign it to the ‘drinkable’ set. This requires us to use a suitable method to determine a threshold such as sampling and testing with a variety of human (or even models of) judges. Given that human beings are all different in their response, we should expect that such a threshold will not be crisp and that some judges will place a particular sample in one set and other judges in another. In a fuzzy approach we allow a fuzzy threshold where a particular sample can belong to both the undrinkable and drinkable sets to some degree by using a membership function such as that shown in Figure 1. This allows us to deal with uncertainty in the measurement device and judgements by people but the problem is to determine the membership functions so that some global optimisation is achieved (e.g. of the average drinkability of the water).

2.2 Type-2 Fuzzy Sets

The fuzzy functions are themselves subject to uncertainty especially when derived from different human judges since we must allow judges to differ in their mem-

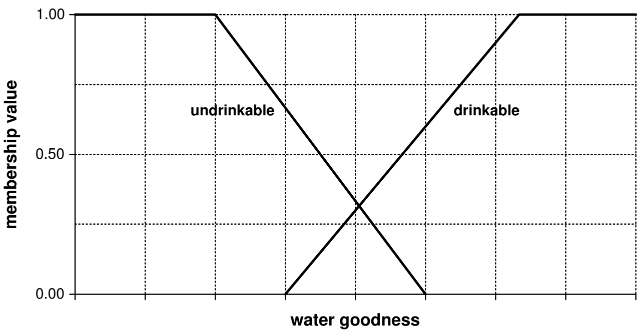

Figure 1: Fuzzy sets for drinkable and undrinkable water

bership values for the same sample. This secondary source of uncertainty can also be accommodated using a fuzzy approach by replacing each membership value by a type-1 fuzzy set. The second order fuzzy set is called a type-2 fuzzy set. For example, for each value of ‘water goodness’ there is now a type-1 fuzzy set which represents the range of membership values associated with it. Set operations of union and intersection have been developed for type-2 fuzzy sets that allow inferencing and higher level processing (composition and defuzzification) to be carried out[22, 17]. Formally a type-2 fuzzy set can be defined in the following way[19]:

Definition 2 *A type-2 fuzzy set is characterised by a fuzzy membership function, i.e. the membership value (or membership grade) for each element of this set is a fuzzy set in [0,1], unlike a type-1 fuzzy set where the membership grade is a crisp number in [0,1].*

2.3 Defuzzification

Once the fuzzy reasoning has been completed it is usually necessary to present the output of the reasoning in a human understandable form, through a process termed *defuzzification*. There are two principle classes of defuzzification, *arithmetic defuzzification* and *linguistic approximation*. In arithmetic defuzzification a method is used to extract the single value in the universe of discourse that ‘best’ (in some sense) represents the arbitrarily complex consequent fuzzy set (Figure 2). The two most popular methods of arithmetic defuzzification are the *centre-of-gravity* (*centroid*) algorithm and the *mean-of-maxima* algorithm. For the consequent set $A = \mu_1/\omega_1 + \mu_2/\omega_2 + \cdots + \mu_N/\omega_N$, the centre of gravity algorithm provides a single value by calculating the imaginary balance point of the shape of the membership:

$$x_g = \frac{\sum_{i=1}^{N}(\mu_i \cdot \omega_i)}{\sum_{i=1}^{N}\mu_i} \tag{1}$$

The mean-of-maxima algorithm finds the point in the universe of discourse with maximum membership grade:

$$x_m = \max_i \mu_i \tag{2}$$

and calculates the mean of all the maxima if more than one maximum is found. This approach is typically used in areas of control engineering where some crisp result must be obtained.

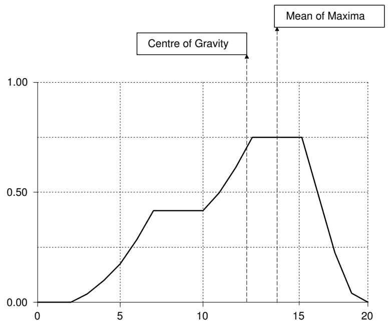

Figure 2: an arbitrary fuzzy output set illustrating the difference between *centre-of-gravity* and *mean-of-maxima* defuzzification

In linguistic approximation the primary terms in the consequent variable's term set are compared against the actual output set in a variety of combinations until the 'best' representation is obtained in natural language. A similarity measure is used to compute the distance between the actual output set and an arbitrary collection of primary terms, connectives and hedges. For example, a shower control variable with primitive terms such as *cold*, *medium* and *hot*, and allowable hedges of *fairly* and *very*, might produce a composite linguistic term such as *medium and fairly hot*. One such similarity metric is the Euclidean distance between fuzzy sets, given by:

$$\delta = \sqrt{\sum_i (\mu_i - \eta_i)^2} \tag{3}$$

where μ_i is the membership of the output set and η_i is the membership grade of the currently considered linguistic approximation — the minimum value of δ will determine the best match. Alternatively, the degree of overlap, γ, of two fuzzy sets, A and B, can be calculated by dividing the area of intersection by the area of union of the sets:

$$\gamma = \frac{A \cap B}{A \cup B} \tag{4}$$

to give a value between zero for disparate sets and one for coincidental sets — the maximum value of γ will determine the best match. A search is then carried out to find the best match whilst attempting to limit the complexity of the combination of terms, in order to produce comprehensible output. This approach is typically used in expert system advisory applications where human users view the output.

2.4 Fuzzy Logic and Probability

It should be made clear that this introduction is not intended to argue for fuzzy logic and against probability. These methods, although fundamentally different, are related (see [20]) and may be used in a complementary manner if desired to cope with different sources of imprecision and uncertainty. In this paper, we aim to show how fuzzy approaches are suited to the context of our medical applications. We leave it to others to argue a separate case for using probability in the same context. The rest of the paper concerns itself with describing how fuzzy logic has been used successfully in medical applications.

3 Categorical Consistency

It is obvious that people learn by experience and that their ability and need to learn varies. The result is that the categories that people learn to recognise need to be revised after a period of experience. In the worst case, this may produce major re-classification of previous material and hence new categories are not necessarily consistent with the old ones in terms of constituency. The provision of relatively novel measuring devices can force this process to occur in a wide variety of applications. In the medical domain, the development of non-invasive body scanners is such an example. In turn, the new measuring devices evolve and re-classification continues to be necessary as each new generation enables higher resolution and new measurements.

Most diseases progress through different stages and complicate the classification process since a disease at one stage can be confused with a different disease at another stage ([11, 13, 12]). However, there is usually an assumption that disease stages can also be categorised and this may not be a clear process for some classes of diseases or conditions. In the worst case, there may simply be a continuous deterioration which has no clear stages associated with it. In this case, the problem of achieving categorical consistency between different patients is severe. Some diseases conditions only occur in soft tissue where some measurement devices have limited resolution and the conditions of measurement and the observer can affect the classification process. The perceptions of observers differ and, in some cases, within observer variations can also be important. In the worst case, we can have all of these sources of imprecision and uncertainty and we must develop a method for coping with this. In this case study, reported in detail in [10, 14], we will show how modelling the acquisition of experience using fuzzy neural networks helps to understand the within observer classification process. We first describe the context and then the fuzzy approach to representing uncertainty.

3.1 The medical context

There has been a marked increase in the number of people engaging in sport/exercise over the last fifteen years or so and a corresponding number of injuries presented in injury clinics. Clinicians have therefore built up significant expertise in recognising and classifying these injuries and have accumulated evidence to support them such as bone scan radiographs. However, the incidence of exercise-induced lower leg pain at particular regional clinics is relatively low. Even at a specialist centre, the incidence may be only 20 per year or so. Given that these cases can be classified into several main categories, it is by no means certain that every clinician will experience all possible classes of injury over a relatively long period of work. Furthermore, the less obvious classes are likely to be confused since they occur over a period of time where correlated features can be forgotten and the stages of the injury are not distinct. The order and time of occurrence of presentation at the clinic is clearly important for these cases since the experience of the clinician plays a dominant part in determining suitable classes.

Our goal is to provide help to a specific expert clinician in classification of sports injuries by providing a system that shows similar cases and how they have been previously classified This is not an easy problem because of the large sources of uncertainty. The system is based on extraction of features from bone scan images that are then submitted to a suitable clustering algorithm (Adaptive Resonance Theory (ART) [2]). It would not be possible with the paucity and imprecision of the data and the state of knowledge of the expert in terms of categorical consistency to provide an ‘automatic’ classification, by, for example, applying standard methodologies for developing expert systems based on rules or back propagation neural networks. This research set out to address whether, with the ART fuzzy neuro-clustering technique, some insights may be provided to the expert that they can use along with their experience and knowledge in order to ensure categorical consistency. It is envisaged that this process is an essential first step in a methodology designed to improve knowledge acquisition for possible future classification systems.

3.2 Uncertainty and Imprecision in Knowledge and Data Sources

There are distinct sources of uncertainty in images, which arise from various causes. Bone is a living tissue and as such is constantly being remodelled. In preliminary research that attempted to extract features using image processing algorithms, a major source of variation in the images was found from normal human variation in leg anatomy. The distance between the knee and ankle could vary as much as 30% over the range of age, sex and other characteristics of the sample population. Bone scanning is a technique that produces an image of the bone on a scintogram and indicates areas of abnormal bone turnover, which may indicate a pathological process. Unlike X-radiographs, scintograms do not reveal distinct clear images, which necessarily directly correspond with anatomical features. Thus the interpretation of the image is not trivial in the majority of cases. For instance, the expert’s judgement of the distinctiveness of a line on the image or whether a line is ‘much longer’ than its width are important in assisting with the classification.

This subjective judgement is clearly a part of the perceptual expertise of the observer, which takes into account the differences between patients, and yet this knowledge is difficult or impossible to effectively capture in numerical feature extraction based systems. Fuzzy sets are based on the principal notion of members belonging to a set to some degree and are thus well matched to modelling observer perceptions of vague information. Knowledge acquisition is notoriously difficult for fuzzy systems since it is particularly difficult to determine membership functions. For this application the uncertain or imprecise nature of the knowledge needs to be captured as input to the clustering neural network (ART). The knowledge acquisition was carried out by interview. On interviewing the expert, it was agreed that the most effective way of capturing the expert's description of the image would be via a questionnaire. After discussion a pro forma was generated that allowed the medical domain expert to view images and record the results. It was decided in the main to use feature attributes which could then be converted into a numerical format. An example of a question about a line on a particular leg on an image is : What best describes the length in relation to the width?

- Much longer
- Longer
- Same
- Less
- Much less

The questionnaire was piloted, some amendments made and finally 203 expert selected images covering approximately 10 years experience were analysed using the questionnaire. Images were selected by the expert by sampling each years images for a selection of all classes including those which were not classifiable. A sample of these questionnaires was double checked by an independent expert. These results were post processed into normalised numeric fuzzy vectors suitable for clustering with a neural network. Since the expert usually viewed both legs on a single scintogram, we collected information about both the left and right legs in our questionnaire.

3.3 Data and Classification analysis

There is no precise way, when viewing an image, of determining the problem with any given shin. Thus, at the time of acquisition, it may only be possible to state that an image is not simply classifiable in terms of the existing know assumed classes. We may assign this image to a class of 'unknown' (denoted by '?'). An image which is classed as '?' does not imply that it cannot be classified at some later time in different circumstances, for example, when a set of similar images are viewed together. 2 data sets were available for clustering and comparison. Each consisted of 203 vectors. An additional file of classes was available whose entries corresponded to the expert's initial classification of each vector. 82 of the vectors classed as 'normal' were all zeros and were removed from the data as their classification is trivial. Thus, there are six classes used in the analysis; Medial

Tibial Syndrome (‘mts’), Normal (‘n’), Stress (‘s’), Healing Stress Fracture (‘hsf’), Patchy(‘p’) and Unknown (‘?’). The particularly large frequency of unknowns (~30%) in the data reflects the problems experts are facing when presented with relatively small numbers of special cases presented in chronological order over a long period of time. This shows the need for developing a method which helps the expert improve the classification process by presenting a subset of similar images which can be viewed all at the same time. This method should allow inferences to be made about the classes of the unknown images even though there is a relatively small amount of data (121 cases) covering several classes (6) with a highly skewed distribution.

The majority of statistical methods (e.g. linear modelling) and some neural methods (e.g. multi-layer feedforward networks) require considerable amounts of well formed statistically sound data for reliable use which is not available in this context. We therefore use a simple clustering method using Fuzzy Neural Networks which will enable coarse judgements to be made about the class of unknowns without statistical support (e.g. levels of significance or certainty). The method is to perform classification analysis of exercise-induced lower leg pain by applying competitive neural network clustering and mapping techniques to type-1 and type-2 fuzzy descriptions of bone scan images of the tibia. The clusters are described and compared with each other and with the experts known classes that would be expected from medical findings. The raw questionnaire data is normalised by considering the range of the linguistic categories of each attribute. These numbers are then considered to be equivalent to fuzzy set membership values for each attribute and are thus type-1 fuzzy data. However, given that some attributes are clearly binary, these vectors may more realistically called hybrids of data types or fuzzy ‘granules’.

Ratio and length are considered by the expert to be important attributes for a large subset of the data, and it was decided to try to capture the 2nd order levels of uncertainty in a representation of these factors. The basic descriptors represented the best estimate that the expert’s interpretation could produce. Therefore, an experimental approach was adopted which involves converting the linguistic categories of the ratio of the sizes of the abnormalities on left and right legs and right and left relative lengths into type-2 fuzzy data. Type-2 fuzzy sets capture a higher level of imprecision than ‘traditional’ (type-1) fuzzy sets by allowing for fuzzy membership grades. Instead of a number in [0,1] the membership grade is a fuzzy set. In other words we can represent the grades as linguistic terms or fuzzy numbers. This is a very powerful notion in that the usual ‘crispness’ required when describing a membership function is removed. The use of type-2 fuzzy sets is on the increase[18, 23]

The membership grades (which are type-1 fuzzy sets) for the type-2 sets for ratio and length are shown in Figure 3. Triangular membership functions have been selected. These are represented as inputs to the networks by the 3 intersection points of the triangle.

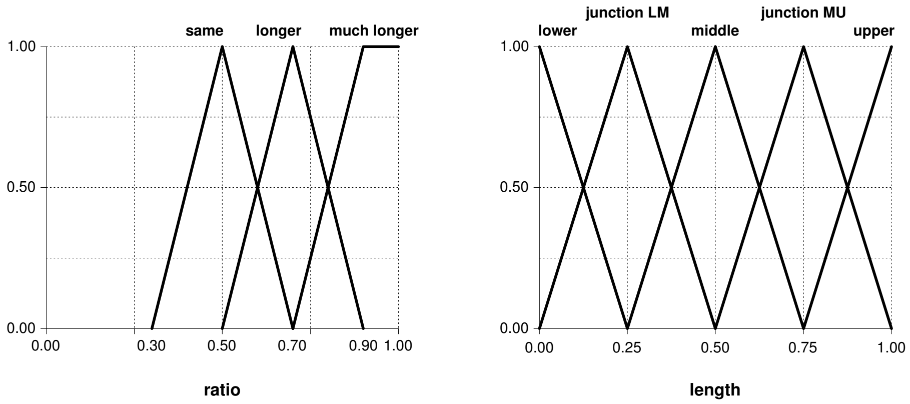

Figure 3: Ratio and length as type-2 fuzzy sets

3.4 Clustering Data Using Fuzzy Neural Networks

There are many approaches which could be taken to cluster fuzzy data. We consider here two simplified competitive fuzzy neural models based on a simple search and match process. One network which was chosen in this research for cluster discovery is based on Adaptive Resonance Theory (ART) [2] since it provides suitable plasticity and does not need large quantities of data. ART networks take feature vectors in the input layer where, after suitable search criteria have been met (category choice), they are individually mapped into a single node in the recognition layer corresponding to a cluster identifier. If a node cannot be mapped into (fails the match criterion) the search continues until all existing nodes have been exhausted, then a new node is established. In this way the ART network is ‘plastic’ though stable. This is a big advantage over fixed topology networks such as the Kohonen and multi-layer feedforward (MLFF) nets for applications such as ours where we wish to discover an unknown number of clusters. The ART algorithm we used and the base program which implements it is presented in [29]. A detailed analysis of the non-linear differential equations that describe ART can be found in[3]. ART can be easily modified to process fuzzy information (type-1 FUZZYART). We wished to modify FUZZYART so that clustering would properly take account of the differences between the information contained in the fuzzy granule description. In order to do this we had to consider both the differences in representation of type-1 and type-2 fuzzy information, as well as how they should be processed. The discovered clusters provide training sets for supervised learning by an ARTMAP neural network[4]. These were used to classify the previously unclassified images and hence improve the classification process. We carried out a series of experiments using the data and the clustering algorithms which are reported in detail in [10] and [14] The experiments involved adjusting the clustering parameter of ART, the ordering of the data and type-1 and type-2 representations. We present a brief selective summary of these results below.

3.5 Results

When neuro-fuzzy clustering is used with with type-1 data presented in the order in which the patients arrived, it was clear that the expert finds the presentation of images in similar groups interesting and causes some review of existing categorisation. However, it was equally clear that, in some cases, a previously categorised image has become less certain. This shows the inherent instability in the classification process due to the paucity of data that also contains uncertainty. The result of this simple experiment was that a set of clusters could be used to guide an expert in reviewing the data base. In assessing the results, the human expert was given the instruction to use the following rules :

1. if the system predictions all agree on a particular prediction category use one of the following terms to describe your opinion of your agreement with the system:
    - very likely (vl)
    - likely (l)
    - possible (po)
    - impossible (ipo)
2. if the system produced differing predictions indicate your preferred prediction category

Of the 38 unknowns, 22 were predicted with agreement from neural models (2 as normal, 9 as healing stress fracture and 11 as MTS). These included the optimistic predictions from clusters with one class. Of the 22 predictions, 5 were not agreed with by the human expert. Of the 16 unknowns predicted without agreement, only 4 were agreed with by the human expert as being unpredictable in the existing classification scheme.

The human expert was, however, able to indicate for each of these cases which classes were not possible for that image. For these unclassifiable images, we may conclude that either the classes currently defined are not exhaustive, or the images were too uncertain in their content and could belong to any of several classes. This represents only ~2% of the total data base which would indicate the latter interpretation is more likely.

We now presented the clustering algorithm with randomly ordered data and repeated the clustering experiments. In addition to the order of input, we wished to explore the parameters which were important for further work. In particular, we wanted to discover if we could improve the prediction of classes of unknowns by using type-2 fuzzy sets rather than type-1.

We found that the majority of clustering results are usefully agreeing with the expert in the classification of unknowns. Inspection of the proportions of agreement within the classes shows that the main sources of agreement are for 'mts'

(80% of 15 cases) followed by 'h'(40% of 15 cases). The 'p' and 's' classes are the most unreliable possibly because of the small number of cases (2). This would be expected by the combination of the specificity of the methods and the paucity and fuzziness of data associated with each class. In the first experiments, the human expert identified 4 unknowns which could not be classified but which could be excluded as belonging to particular classes. These impossible classes were predicted by some neural models (ART clustering). No neural model was able to match the human expert. Deductions from this are limited.

The data indicates that type-1 produces generally more unclassifiable cases than type-2. This would be expected from the usual trade off between false and true positives when a decision criterion has been consistent in a neural method. Overall, the accuracy of single neural models is at best 70% and at worst, 27%.

3.6 Conclusion

This section has discussed work on using a fuzzy approach to assist the classification of radiographic images. Experiments revealed that care had to be taken with the plasticity of the neural models if robust conclusions are to be drawn. However, by simply providing clusters of related images to a human expert, he was able to eliminate the majority of previously unclassified images by providing a class and its likelihood for an image. The sets of clusters generated by the neural models are dependent on the order of presentation of the data and the setting of control parameters. This appears to be a good model of knowledge acquisition by experts in this domain. The overall conclusion is that the use of fuzzy representation and the fuzzy neural clustering methods has improved the classification process of the shin images in helping achieve categorical consistency despite the paucity of data and its inherent uncertainty. The work has also shown that it is possible to automatically classify new images by using clustering neural networks combined with a suitable interface and automatic feature extraction software.

4 Diagnostic Monitoring

There are many medical devices available to measure some physiological parameter(s) of a patient. Such measured (numerical) parameters are usually recorded and then interpreted in context by a clinician to monitor the clinical status of the patient. This physiological monitoring is then most often combined with other patient information to form a diagnosis of their condition. The process of interpreting these parameters in context using clinical knowledge can be replicated by expert systems. However, any physical measurement device has inherent uncertainty (inaccuracy and imprecision) in its results, and there is always additional uncertainty in the contextual data and vagueness in the clinical knowledge. Fuzzy logic provides a natural framework for representing and incorporating all these sources of uncertainty into an expert system for diagnostic monitoring. We present a case study of such a fuzzy expert system in the field of immediate neonatal assessment at birth.

4.1 Neonatal Assessment at Birth

Childbirth is a stressful experience for both mother and infant. Even during normal labour every infant is being regularly deprived of oxygen as maternal contractions, which increase in frequency and duration throughout labour until delivery, restrict blood supply to the placenta. This oxygen deprivation can lead to fetal ‘distress’, permanent brain damage and, in the extreme, fetal death.

In the 1950’s Virginia Apgar introduced a scoring system [1] to indicate the requirement for neonatal resuscitation, which was originally intended to be assessed objectively by an independent observer. It has since been adopted almost universally in the developed world. However, it is now widely assigned by the attending clinician, is therefore subjective and is often assigned retrospectively. The Apgar score is also affected by factors that existed prior to the onset of labour such as congenital abnormalities of the fetus and events immediately after delivery. The need remains to establish an objective, physiologically based, immediate assessment of the health of the newborn infant which can be used to accurately evaluate obstetric care. Such an assessment could be used to guide neonatal care, provide individual feedback to clinicians, audit overall hospital performance, teach inexperienced clinicians and assess the impact of new technologies.

An assessment of neonatal outcome may be obtained from analysis of blood in the umbilical cord of an infant immediately after delivery, and has been recommended in the United Kingdom by the Royal College of Obstetricians and Gynaecologists [24]. 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 (pO_2) and low partial pressure of carbon dioxide (pCO_2). Oxygen in the blood fuels *aerobic* cell metabolism, with carbon dioxide produced as ‘waste’. Thus the blood returning from the fetus has 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). 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.

A sample of blood is taken from each of the blood vessels in the clamped umbilical cord and a blood gas analysis machine measures the pH, pO_2 and pCO_2. A parameter termed *base deficit of extracellular fluid* (BD_{ecf}) can be derived from the pH and pCO_2 parameters [25]. 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. An interpretation is then made based on the pH and BD_{ecf} parameters (‘the acid-base status’) of both arterial and venous blood.

There are, however, a number of difficulties with such umbilical acid-base analysis. Difficulties in obtaining the samples can result in two samples from the same vessel or mixed samples, whilst blood in the syringe can alter due to exposure to air. Blood gas analysis machines require regular calibration and quality control checks to ensure continuing performance to the manufacturer's specifications. Careful retrospective analysis of the acid-base results obtained during a trial on electronic fetal monitoring highlighted a 25% failure rate to obtain arterial and venous paired samples with all parameters [27]. This sampling error rate is broadly in line with other studies in which the importance of paired samples was recognised. The study also highlighted the fact that considerable expertise was required to reliably recognise these errors and accurately interpret the results.

4.2 A Fuzzy Expert System for the Analysis of Umbilical Acid-Base Status

A fuzzy expert system was developed for the analysis of umbilical cord acid-base status, encapsulating the knowledge of leading obstetricians, neonatologists and physiologists gained over years of acid-base interpretation. The expert system combines knowledge of the errors likely to occur in acid-base measurement, physiological knowledge of plausible results, statistical analysis of a large database of results and clinical experience of acid-base interpretation. It automatically checks for errors in input parameters, identifies the vessel origin (artery or vein) of the results and provides an interpretation in an objective, consistent and intelligent manner.

This process is carried out in two distinct phases; *validation* of parameters and *interpretation* of parameters. The expert system comprises two separate fuzzy rule bases, one for each phase.

4.3 Validation of Parameters

The three measured parameters (pH, $p\mathrm{O}_2$, and $p\mathrm{CO}_2$) for each sample are introduced to the expert system without labelling — the detection of parameter errors and identification of vessel origin for each sample is an entirely automatic data-driven process carried out by the expert system.

There are two main classes of parameter error:

- *sample source* — for some reason the blood drawn into the two syringes does not constitute the intended samples of arterial and venous blood (for example, it is relatively easy to inadvertently stick the needle right through a cord artery and mistakenly draw blood from the vein, due to either inadequate umbilical cord segment or poor sampling technique);

- *parameter inaccuracy* — the measurements reported by the blood gas analysis machine do not accurately represent the *true* parameter values of the blood sample.

This can be caused by either:

- *machine error* — the blood gas analysis machine has drifted somewhat from *true* calibration;
- *sample error* — the sample contains air bubbles or other miscellaneous contaminants;
- *time delays* — the umbilical cord was not clamped immediately, or there is some time delay between taking the samples and then in introducing samples into the blood gas machine for measurement (note that it has been experimentally verified that parameter values remain stable in a clamped cord segment for around one hour).

As a first step, the expert system examines the relationship between the pH and $p\mathrm{CO}_2$ parameters for each sample independently. Briefly, there is a biochemical relationship between these parameters such that, for neonatal blood, bounds can be established on a physiologically possible $p\mathrm{CO}_2$ for any given pH. The expert system checks that the parameters fall within these bounds to exclude, for example, samples where non-blood fluid has been accidentally sampled. Once it has been established that the parameters are compatible with neonatal blood, the parameters can be compared across samples to detect further errors and identify vessel origin. A number of experiments were carried out in order to establish the maximum likely error in each parameter (given a known sample source) as summarised in Table 1. The consequent likely error in the derived base deficit parameter was determined mathematically from the component uncertainties.

	δpH	$\delta p\mathrm{CO}_2$	$\delta p\mathrm{O}_2$
Arterial	0.025	0.61 kPa	0.31 kPa
Venous	0.010	0.24 kPa	0.15 kPa

Table 1: the combined *maximum likely* uncertainty (two standard deviations from the mean) in each umbilical cord acid-base parameter

Physiologically, it can be expected that the arterial pH (pH^{A}) is *lower* than the venous pH (pH^{V}), the arterial $p\mathrm{CO}_2$ ($p\mathrm{CO}_2{}^{\mathrm{A}}$) is *higher* than the venous $p\mathrm{CO}_2$ ($p\mathrm{CO}_2{}^{\mathrm{V}}$), the arterial $p\mathrm{O}_2$ ($p\mathrm{O}_2{}^{\mathrm{A}}$) is *lower* than the venous $p\mathrm{O}_2$ ($p\mathrm{O}_2{}^{\mathrm{V}}$), and the arterial $\mathrm{BD}_{\mathrm{ecf}}$ (BD^{A}) is *higher* than the venous $\mathrm{BD}_{\mathrm{ecf}}$ (BD^{V}). These facts are expressed in the definitions of parameter differences as shown in Table 2 such that the Δ 's are all expected to be positive.

If two good venous samples were obtained, then each parameter should differ by amounts close to, or less than, the venous values shown in Table 1. As two samples may both be accidentally obtained from the vein, both from the arteries, one may be mixed arterial-venous, or both may be mixed, a ‘safe’ vessel identification rule may be that if all parameters differ by more than the largest uncertainties in Table 1, then the samples can definitely be taken as a true arterial-venous pair.

Difference	*Definition*
ΔpH	venous pH − arterial pH
ΔpCO$_2$	arterial pCO$_2$ − venous pCO$_2$
ΔpO$_2$	venous pO$_2$ − arterial pO$_2$
ΔBD$_{\text{ecf}}$	arterial BD$_{\text{ecf}}$ − venous BD$_{\text{ecf}}$

Table 2: the definition of Δ for acid-base parameters

Given that the lowest pH is initially labelled as the artery (or if the pH's are the same, the highest pCO$_2$ is labelled as the artery, or if the pH's and pCO$_2$'s are the same, the lowest pO$_2$ is labelled as the artery), then the list of possible sample differences and their associated vessel identification are shown in Table 3. A '0' indicates that Δ*parameter* is zero, a '−' indicates that Δ*parameter* is negative, and a '+' indicates that Δ*parameter* is positive.

ΔpH	ΔpCO$_2$	ΔpO$_2$	*Origin*
0	0	0	definitely same
0	0	+	probably same
0	+	0	probably same
0	+	−	probably mixed
0	+	+	probably different
+	0	0	probably same
+	0	−	probably mixed
+	0	+	probably different
+	−	0	probably mixed
+	−	−	definitely mixed
+	−	+	definitely mixed
+	+	0	probably different
+	+	−	definitely mixed
+	+	+	definitely different

Table 3: list of possible sample differences and the associated vessel identification

A fuzzy rule-base was designed to produce the target behaviour shown in Table 3, with smooth transitions between each of the categories. The rule-base consisted of a set of five rules relating the differences in fuzzy input parameters (the pH, pCO$_2$, and pO$_2$ in both samples) to a single fuzzy output variable, the *origin* of samples. Each input parameter was first fuzzified to possess a width equal to the largest (arterial) uncertainties in Table 1, as shown for example in Figure 4. The fuzzified input variables were then passed through the vessel identification rule-set with rules of the form:

IF pH^{A} *IS-EQUAL-TO* pH^{V}
AND $p\mathrm{CO}_2{}^{\mathrm{A}}$ *IS-ABOUT-EQUAL-TO* $p\mathrm{CO}_2{}^{\mathrm{V}}$
AND $p\mathrm{O}_2{}^{\mathrm{A}}$ *IS-ABOUT-EQUAL-TO* $p\mathrm{O}_2{}^{\mathrm{V}}$
THEN *origin IS paired*

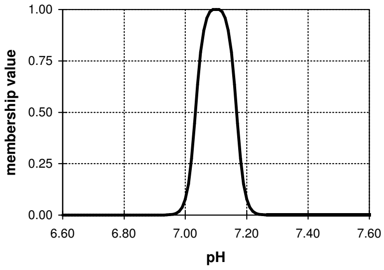

Figure 4: An example of a pH input parameter with a fuzzy width (note that the width has been exaggerated for clarity)

Linguistic approximation of the *origin* output variable was used to determine the appropriate vessel labelling. A linguistic output corresponding to *different*, *mixed/different* or *not same* causes the vessels to be labelled as an arterial-venous pair. Any other linguistic output, or the presence of only one input sample, caused the sample to be labelled as a single venous vessel. An arterial-venous pair would then have its input variables re-initialised with the crisp values of the input parameters, fuzzified to have a width equal to those in Table 1. The $\mathrm{BD}_{\mathrm{ecf}}$ for each vessel is then calculated from the pH and $p\mathrm{CO}_2$ parameters and then fuzzified to the derived width.

If both vessels were missing, both the arterial and venous parameters would be initialised with $\mu_A(x) = 1$ across the universe of discourse. In such a situation all rules fire with maximum strength and the output of all variables tends to $\mu_A(x) = 1$. In practice such a situation is very rare (1 case out of $> 10\,000$), and the much more common occurrence is the single vessel. In this case, a single vessel is always labelled as venous for the reasons that firstly, the vein is *much* easier to sample than the artery and hence an arterial sample without a venous sample is unlikely in the extreme and, secondly, as the artery is effectively ‘worse’ than the vein from a health point of view, assuming a single vessel is a vein is the safest option. It might be thought that the subsequent arterial parameters would simply be initialised with $\mu_A(x) = 1$ across the entire fuzzy set. However, this ignores the fact that, physiologically, the arterial parameters would be such as to maintain positive Δ’s — i.e. if the arterial pH was not known, it could still be assumed to be lower than the venous pH. Thus, the actual procedure was to initialise a missing arterial pH parameter with a fuzzy set consisting of the inverted left-hand edge of the venous fuzzy set, and to initialise a missing $\mathrm{BD}_{\mathrm{ecf}}$ parameter with a fuzzy set consisting of the inverted right-hand edge of the venous fuzzy set. This is demonstrated in

Figure 5, in which missing arterial values have been set to $\mu_A(x) = 1$ relative to venous values of pH = 7.10 and $\text{BD}_{\text{ecf}} = 9.7$ mmol.l^{-1}.

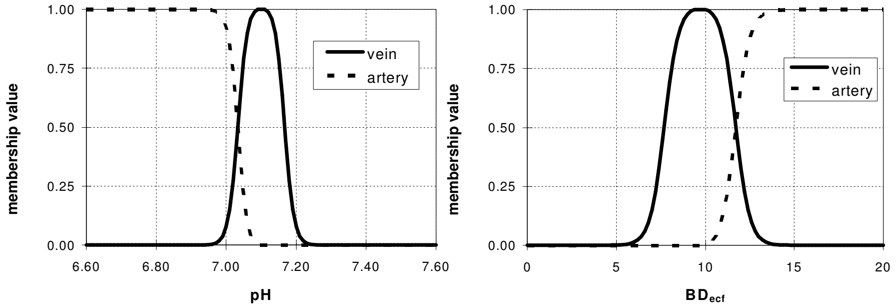

Figure 5: illustration of missing arterial values given known venous values of pH = 7.10 and $\text{BD}_{\text{ecf}} = 9.7$ mmol.l^{-1} (note that parameter widths have been exaggerated for clarity)

4.4 Interpretation of Parameters

Once vessel identification has been carried out, the sample(s) are passed through the interpretation rules. The basic principles of acid-base analysis elicited from the experts were that:

(i) *acidemia* is based on the absolute value of arterial pH (lower arterial pH implies worse *acidemia*), refined by the value of the venous pH;
(ii) *component* is based on arterial BD_{ecf} (high BD_{ecf} implies *metabolic* component, low BD_{ecf} implies *respiratory* component), refined by venous BD_{ecf}; and
(iii) *duration* is based on pH and BD_{ecf} differences (smaller differences imply *chronic* duration, larger differences imply *acute* duration), refined by absolute arterial values.

These basic principles were encapsulated in a set of fuzzy rules developed over a series of elicitation and comparison sessions with acknowledged umbilical acid-base experts. Three fuzzy output variables (*acidemia*, *component*, and *duration*) were utilised in rule consequences, with the availability of graphical output of the consequence fuzzy sets. The consequent fuzzy sets were defuzzified by both the *centroid* method and linguistic approximation, and the results were validated against international expert opinion.

4.5 Validation of the Fuzzy Interpretation

The cases for each task were selected by the independent engineer from the database of over 10 000 results (approximately 400 abnormals), but this selection presented serious problems. Cases could not be selected from the entire database on a uniform random basis, as this would have resulted in approximately 75% paired arterial-venous samples, and approximately 98% *normal* interpretations. In essence it was desired to uniformly span the *target* outputs, so that a roughly even spread across the various output sets would have been obtained from the combined experts (and expert system). However, this pre-supposed that the output was known — which it obviously wasn't for the validation study. An in-house expert could have been used to select difficult and/or representative cases, but due to the restricted number of experts available this was not feasible. The problem was solved by using a categorisation already obtained by a previously developed crisp expert system to guide the selection of cases. Two sets of fifty cases were randomly selected to roughly span the crisp expert system categorisations. This ensured that a few cases were obtained from a variety of conditions, including results that had parameter errors, results from a single vessel, and results ranging from metabolic acidemia to normal.

The centroids of the fuzzy output variables were combined into a single index by:

$$condition = acidemia + \frac{component}{20} + \frac{duration}{10} \tag{5}$$

where the relative weighting of the three terms had been determined empirically. Given that the three output variables are arranged in such a way that low scores indicate a worsening condition for the infant, to the extreme *severe*, *metabolic*, *chronic acidemia*, this index can be thought of as indicating the *health* of the infant as represented by its acid-base balance at birth. The experts were asked to rank fifty cases from 'worst' to 'best', in terms of likelihood that the infant may have suffered damage during labour, on the basis of the acid-base information alone. Spearman rank order correlation was used to determine the degree of association between the expert system's ranking of cases, specified by the index described above, and the experts' ordering. Note that this is effectively the same as minimising the mean square error between the desired rankings and the obtained rankings.

The experts were given the two sets of pH and $\mathrm{BD}_{\mathrm{ecf}}$ parameters from each of fifty cases, and were asked to indicate their opinion of the closest linguistic interpretation for three linguistic variables; *acidemia*, *component*, and *duration*. For each variable they were instructed to mark *zero*, *one* or *two* terms to indicate the closest match. This was specifically designed to allow the expert to mark two adjacent labels if they felt a result fell in-between two labels, or to mark no label if there was insufficient information, or no label was appropriate. To measure the agreement between two expert's linguistic categorisation a measure of (nominal) categorical agreement was required. The kappa statistic [5] was used to measure exact agreement between experts and the expert system linguistic outputs and weighted kappa [6] was used for partial agreement.

4.6 Results

The individual inter-expert and expert-*Fuzzy* Spearman rank order correlation coefficients obtained are shown in Table 4. The average inter-expert agreement is calculated by taking the average of each expert against the other *three* experts, and the average *Fuzzy* agreement by taking the average of agreement with all *four* experts. As can be seen, the fuzzy expert system performed exceptionally well against experts *A*, *B*, and *C*. These three experts had taken place in the previous study, and the average expert system agreement with these three is 0.94 — slightly lower correlation was obtained against expert *D*, although the fuzzy expert system was no worse than the other experts. These results are illustrated in Figure 6, in which each of the expert's rankings are plotted against the fuzzy expert system rankings — perfect agreement would result in a diagonal line from (1,1) to (50,50).

The results of the linguistic interpretation agreement were generally found to be relatively low, even for weighted kappa, both for inter-expert agreement and expert-*Fuzzy* agreement. An attempt was made to investigate the effect of different pH and BD_{ecf} weights on these linguistic agreements, but in general it was found that performance was not significantly increased above the results achieved with default weights. It would seem that while the experts can agree on relative 'badness' of results, placing a linguistic label on the results is much more subjective.

4.7 Conclusion

The resultant fuzzy expert system explicitly represented uncertainty in both the input data and the knowledge base. Although presented as a single achievement, the eventual fuzzy system was only arrived at after a long, iterative development process, beginning with the creation of a crisp expert system [9, 8], followed by an intermediate fuzzy expert system, which performed only interpretation of previously validated parameters [7]. The fuzzy expert system was tested in a validation study and was found to perform favourably compared to internationally acknowledged domain experts.

Table 4: Agreement for numeric interpretation by rank order correlation

Expert	A	B	C	D	*Fuzzy*
A	—	0.899	0.888	0.577	0.950
B	0.899	—	0.908	0.701	0.931
C	0.888	0.908	—	0.537	0.925
D	0.577	0.701	0.537	—	0.606
Average	0.788	0.836	0.777	0.605	0.853

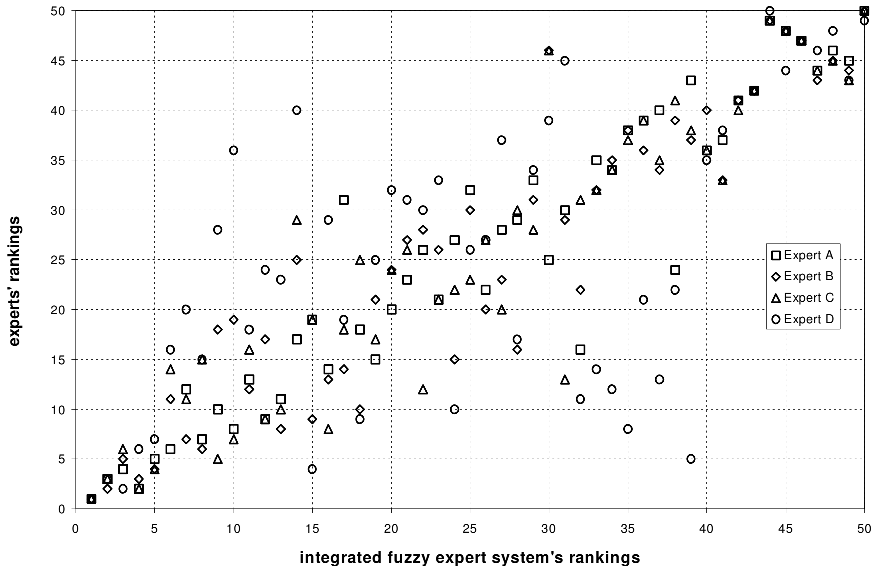

Figure 6: Graph of four experts' rankings against the integrated fuzzy expert system

5 Scoring

A simple approach to many problems of classification is to use some sort of 'scoring' approach. Perhaps the most well known is credit scoring where a persons credit worthiness (for a loan for example) is arrived at by attaching numeric scores to a number of different aspects of the individual (e.g. employment status) and adding the scores up (perhaps with some weighting).

We have used type-2 fuzzy relations to capture nursing perceptions in nursing assessment as an alternative to a numerical scoring approach. The detail of the work is published in [16] and (from a nursing perspective)[21]

5.1 The medical context

The clinical reality of nursing requires nurses to make decisions arising from an ongoing holistic assessment of the patient need for nursing care, based on an extensive range of knowledge. Nurses concurrently assess patient need for nursing care in several domains of concern to nursing before deciding where the primary focus of nursing attention should be directed. Holistic nursing assessment takes a number of environmental domains of patient need into account, as well as the need for clinical intervention. For the purposes here we simplify the framework for assessment to five domains.

1. Nursing is often carried out in conjunction with medical diagnosis and treat-

ment. A primary focus of nursing concern is the physical/medical condition or diagnosis of the person. Therefore the initial domain of assessment is named ‘Physical/Medical Condition’.

2. A domain named ‘Complicating Factors’, which may affect the initial condition, is then also considered and taken into account.
3. The physical capability or ‘Dependency’ domain of the individual is always assessed.
4. The patient’s ability to understand and co-operate with suggested interventions and the support available to the person from their family and environment will also affect the amount and type of nursing intervention that will be provided. This domain has been summarised as the ‘Psycho-Social’ domain.
5. The requirement for the more obvious array of nursing clinical interventions, as titrated to patient need and condition, are combined in the ‘Clinical Intervention’ domain.

It is suggested that these five domains provide the context of the patient need for nursing care and intervention. These ‘top level’ domains are clearly imprecise and subjective and are difficult, if not impossible to measure. Added to this imprecision, there is, within each domain, a degree, or priority, of need to be determined before any required intervention is applied.

To demonstrate the translation of such a framework into Fuzzy Logic a hypothetical patient requiring elective surgery is considered. The assessments are made over a period of four to five days during the peri-operative and post-operative recovery, and a summary of two separate daily assessments made. The patient is a frail 48yr old, admitted for relief of symptoms of cancer by the insertion of a celestin tube under anaesthetic to enable parenteral feeding. On return from theatre, his post-operative Physical/Medical condition is assessed as ‘potentially unstable’, requiring routine regular observations. Complicating Factors of nausea and pain/comfort are also assessed as ‘potentially unstable’, also requiring regular observation. His requirements for Clinical Intervention are assessed as ‘complex’, as antibiotics, intravenous analgesia and anti-emetics are administered as necessary. His Dependency needs are assessed as ‘heavily dependent’, both due to his generally frail physical condition, plus the necessity for added assistance in the post-operative recovery phase. His need for emotional/educational support in the Psycho-Social domain is assessed as ‘moderate’, somewhat eased by post-operative sedation. This denotes a ‘more acutely unwell’ patient in the context of surgery on a person with cancer and secondary spread, who requires frequent nursing assessment, clinical intervention and physical support in the post-operative period. A summary might be: “Frail, requiring frequent intervention for relief of symptoms.”

Both the domains and the degree of need within each domain are imprecise, using nursing perceptions to define or explain the nursing assessment. The imprecision in the linguistic perceptions granulates the complexity of patient need into recognisable fragments of knowledge. To simplify this to the summation of numerical scores requires translation from linguistic terms to numbers. Our modelling of this problem removes the need for translation.

5.2 Type-2 Relations and Nursing Assessment

A definition of a type-2 fuzzy relation is as follows:

Definition 3 *A type-2 fuzzy relation is a type-2 fuzzy set defined on the Cartesian product of the crisp sets* $X_1, X_2, \ldots, X_n$ *where the tuples* $(x_1, x_2, \ldots x_n)$ *have varying degrees of membership which are type-1 fuzzy sets.*

In other words, the type-2 fuzzy relation indicates a degree of membership which is itself a type-1 fuzzy set - not a number in [0,1]. In our work we use type-2 relations to modelling the nursing paradigm. The mathematical detail can be found in John and Lake(2001a).

The approach is to model the five nursing domains using type-2 fuzzy sets. The expert nurse provided the diagrams like that in Figure 7 to show how the words could be represented on a linear domain.

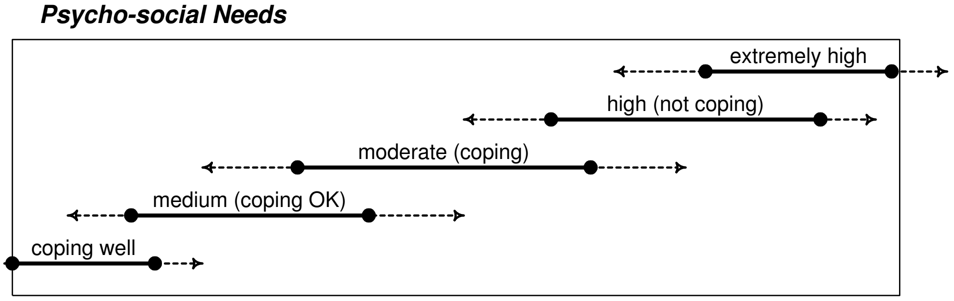

Figure 7: The Psycho-Social Domain

These diagrams were translated into type-2 fuzzy sets. For example we have ‘Dependency’ which would have possible membership grades (type-1 fuzzy sets) ‘independent’, ‘becoming independent’, ‘dependent’, ‘heavily dependent’ and ‘totally dependent’. These membership grades have to be determined somehow. In this instance nurses who were expert in nursing assessment provided ranges for coverage of the words. Gaussian membership functions were used to represent the type-2 membership grades. Figures 8 and 9 show the representation of the type-2 fuzzy sets Physical/Medical and Complicating Factors.

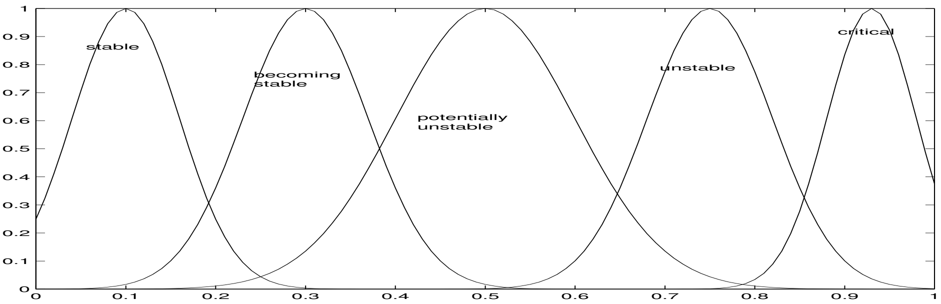

Figure 8: The Type-2 Fuzzy Set Physical/Medical

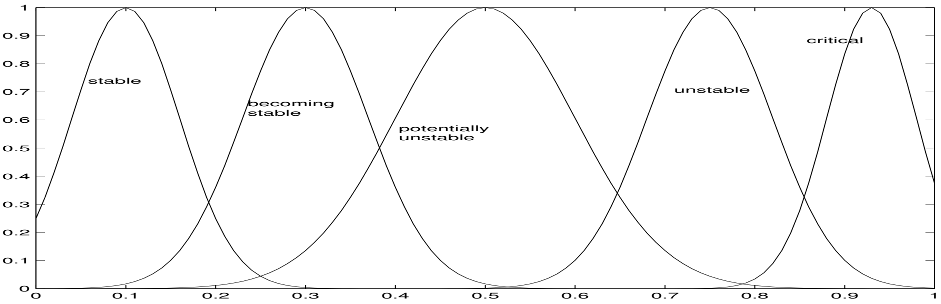

Figure 9: The Type-2 Fuzzy Set Complicating Factors

These Type-2 fuzzy sets for the five domains can then be combined using type-2 relations to produce a type-2 set that we label 'Overall Condition'. The nursing understanding of the range of patient acuity for overall condition from 'stable' to 'critical', is also imprecise. However, nursing assessment pays attention to all aspects or domains of the patient's well being to create the summary. Let us consider a patient with membership in the type-2 sets as potentially unstable in the Physical/Medical domain, becoming unstable in Complicating Factors, complex in Clinical Intervention, heavily dependent in dependency and moderate needs in Psycho-Social. The overall condition is described as unstable. For this particular patient we carry out the 'meet'[1]. Figure 10 provides the resulting membership grade of this particular patient in the type-2 fuzzy set 'Overall Condition'.

Suppose we now have a patient who has different membership in each type-2 fuzzy set - critical in Physical/Medical, unstable (Complicating Factors), critical(Clinical Interventions), totally dependent (Dependency) and moderate (Psycho-social needs). In this case we get the result in Figure 11.

The overall condition is described as critically ill

The shape of the membership grades for each patient is quite different. The ordering is intuitively correct although one would perhaps expect the critically ill membership grade to be further along the axis.

---

[1] The software used is that provide by Professor Mendel at http://sipi.usc.edu/m̃endel/.

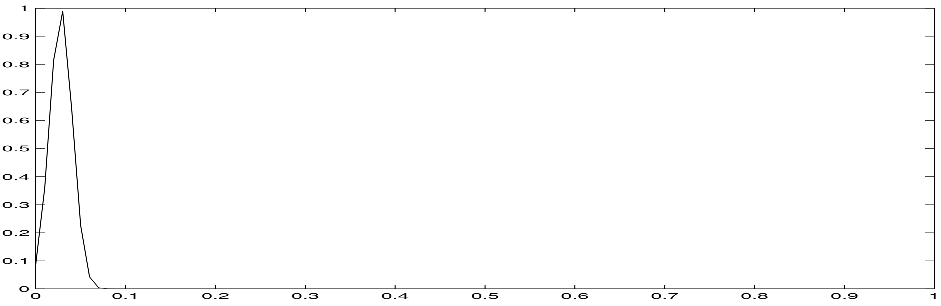

Figure 10: The Membership of the First Patient in Overall Condition

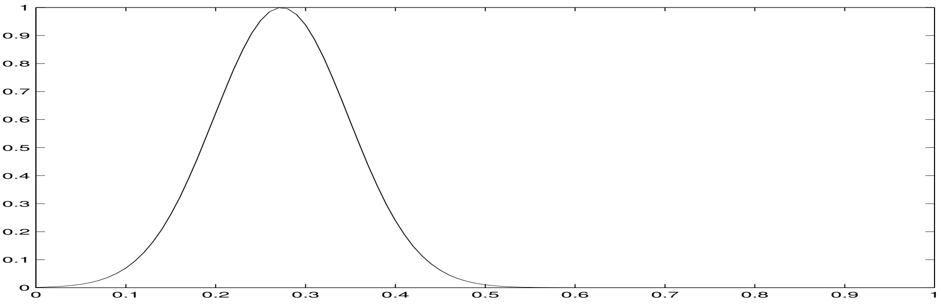

Figure 11: The Membership of the Second Patient in Overall Condition

5.3 Conclusion

It appears that type-2 relations have much to offer in nursing assessment. This work is not as mature as that discussed in the previous sections but nevertheless offers an insight into the role that type-2 fuzzy sets and relations can play in medical diagnosis. In particular we would argue that this approach offers a more intuitive approach than scoring systems especially in an inherently imprecise domain such as nursing

6 Summary

The use of fuzzy approaches has been successful in our work so far, which has addressed particularly difficult problems in the medical field involving classification and perception by experts of uncertain measured parameters, and visual and linguistic information. We see future directions developing fundamental methods such as supervised learning of type-2 (linguistic) fuzzy sets and exploring their applicability in the very rich and important area of medical diagnosis and analysis.

7 Acknowledgements

Dr. Innocent would like to thank Dr. Barnes of Leicester General Hospital who provided the expertise for the shin image analysis. Dr. Garibaldi would like to thank Prof. Emmanuel Ifeachor for his contribution to this work, and to the clinical experts whose help was invaluable in the development and validation of the umbilical acid-base expert system: Prof. Karl Rosén, Prof. James Low, Dr. Jenny Westgate, Dr. Mark Davies, Dr. Sarah Beckley and Mrs. Maureen Harris. The work on using type-2 fuzzy sets to model nursing perceptions is collaborative with Sarah Lake of Nelson Hospital, New Zealand.

References

[1] V. Apgar. A proposal for a new method of evaluation of the newborn infant. *Current Researches in Anesthesia and Analgesia*, 32:260–267, 1953.

[2] G.A. Carpenter and S Grossberg. Art 2: Selforganisation of stable category recognition codes for analog input patterns. *Applied Optics*, 26:4919–4930, 1987.

[3] G.A. Carpenter and S. Grossberg. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, And Image Processing*, 37:54–115, 1987.

[4] G.A. Carpenter, S. Grossberg, N. Markuzan, J.H. Reynolds, and D.B. Rosen. Fuzzy ARTMAP: A Neural Network Architecture For Incremental Supervised Learning Of Analog Multidimensional Maps . *IEEE Trans. On Neural Networks*, 3(5):698–712, 1992.

[5] J. Cohen. A coefficient of agreement for nominal scales. *Educational Psychological Measurement*, 20:37–46, 1960.

[6] J. Cohen. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70:213–220, 1968.

[7] J.M. Garibaldi and E.C. Ifeachor. Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation. *IEEE Transactions Fuzzy Systems*, 7(1):72–84, 1999.

[8] J.M. Garibaldi, J.A. Westgate, and E.C. Ifeachor. The evaluation of an expert system for the analysis of umbilical cord blood. *Artificial Intelligence Medicine*, 17(2):109–130, 1999.

[9] 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. *Artificial Intelligence Medicine*, 10(2):129–144, 1997.

[10] P. Innocent, M. Barnes, and R.I. John. Application of the Fuzzy ART/MAP and MinMax/MAP Neural Network models to Radiographic Image Classification. *Artificial Intelligence in Medicine*, 11(3):241–263, 1997.

[11] P.R. Innocent. Clinical diagnosis using fcm and fuzzy temporal reasoning. *Conference Proceedings SCB99 (ICSC Rochester. N.Y.)*, pages 490–496, 1999.

[12] P.R. Innocent. Fuzzy symptoms and a decision support index for the early diagnosis of confusable diseases. *Advances in Soft Computing*, pages 79–91, 2000.

[13] P.R. Innocent. A lightweight fuzzy process to support early diagnosis of confusable diseases using causation and time relationships. *FUZZ-IEEE2000*, 2:516–521, 2000.

[14] P.R. Innocent, R.I. John, and M.R. Barnes. Neuro-fuzzy Models of Radiographic Image Classification. In P.S. Szczepaniak, P.J.G. Lisboa, and J. Kacprzyk, editors, *Fuzzy Systems in Medicine*, pages 361–393. Springer-Verlag, 2000.

[15] J.S.R. Jang, C.T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing*. Prentice Hall, 1997.

[16] R. John and S. Lake. Modelling nursing perceptions using type-2 fuzzy sets. *Proceedings of EUROFUSE 2001 Workshop on Preference Modelling and Applications*, pages 241–246, 2001.

[17] R.I. John. Type 2 Fuzzy Sets: An Appraisal of Theory and Applications. *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems*, 6(6):563–576, 1998.

[18] R.I. John. Type 2 Fuzzy Sets for Knowledge Representation and Inferencing. *Proc. 7th Intl. Conf. on Fuzzy Systems FUZZIEEE'98*, pages 1003–1008, 1998.

[19] N.N. Karnik and J.M. Mendel. Introduction to Type-2 Fuzzy Logic Systems. In *Proc. 7th Intl. Conf. on Fuzzy Systems FUZZ-IEEE'98*, pages 915–920, 1998.

[20] B. Kosko. *Fuzzy Engineering*. Prentice-Hall International, 1997.

[21] S. Lake and R. I. John. Patient Assessment in Nursing Care using Fuzzy Logic. *Nursing Informatics Congress 2000, New Zealand*, pages 468–475, 2000.

[22] Q. Liang and J.M. Mendel. An Introduction to Type-2 TSK Fuzzy Logic Systems. In *Proc. 8th Intl. Conf. on Fuzzy Systems FUZZ-IEEE'99*, pages 1534–1539, 1999.

[23] Mendel J. M. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice-Hall, Upper Saddle River, NJ., 2001.

[24] Royal College of Obstetricians and Gynaecologists. Recommendations Arising from the 26th RCOG Study Group. In J.A.D. Spencer and R.H.T. Ward, editors, *Intrapartum Fetal Surveillance*, page 392. RCOG Press, London, 1993.

[25] O. Siggaard-Andersen. *The Acid-Base Status of the Blood (4th edn)*. Munksgaard, Copenhagen, 1976.

[26] F. Steimann and K.P. Adlassnig. Fuzzy medical diagnosis, http://citeseer.nj.nec.com/160037.html, 2001.

[27] J.A. Westgate, J.M. Garibaldi, and K.R. Greene. Umbilical cord blood gas analysis at delivery: A time for quality data. *British Journal Obstetrics Gynaecology*, 101:1054–1063, 1994.

[28] J. Winer and S. Sandri. Fuzzy temporal/categorical information in diagnosis. *Journal of Intelligent Information Systems*, 13(1–2):9–26, 1999.

[29] J.M. Zurada. *Introduction To Artificial Neural Systems*. St. Paul, Mn : West Publishing Company, 1992.