Expert system support using Bayesian belief networks in the diagnosis of fine needle aspiration biopsy specimens of the breast

P W Hamilton, N Anderson, P H Bartels, D Thompson

Abstract

Aim—To develop an expert system model for the diagnosis of fine needle aspiration cytology (FNAC) of the breast.

Methods—Knowledge and uncertainty were represented in the form of a Bayesian belief network which permitted the combination of diagnostic evidence in a cumulative manner and provided a final probability for the possible diagnostic outcomes. The network comprised 10 cytological features (evidence nodes), each independently linked to the diagnosis (decision node) by a conditional probability matrix. The system was designed to be interactive in that the cytopathologist entered evidence into the network in the form of likelihood ratios for the outcomes at each evidence node.

Results—The efficiency of the network was tested on a series of 40 breast FNAC specimens. The highest diagnostic probability provided by the network agreed with the cytopathologists' diagnosis in 100% of cases for the assessment of discrete, benign, and malignant aspirates. Atypical probably benign cases were given probabilities in favour of a benign diagnosis. Suspicious cases tended to have similar probabilities for both diagnostic outcomes and so, correctly, could not be assigned as benign or malignant. A closer examination of cumulative belief graphs for the diagnostic sequence of each case provided insight into the diagnostic process, and quantitative data which improved the identification of suspicious cases.

Conclusion—The further development of such a system will have three important roles in breast cytodiagnosis: (1) to aid the cytologist in making a more consistent and objective diagnosis; (2) to provide a teaching tool on breast cytological diagnosis for the non-expert; and (3) it is the first stage in the development of a system capable of automated diagnosis through the use of expert system machine vision.

Department of Pathology, The Queen's University of Belfast, Northern Ireland
P W Hamilton
N Anderson

Optical Sciences Centre, University of Arizona, Tucson, Arizona, USA
P H Bartels
D Thompson

Correspondence to: Dr Peter W Hamilton Quantitative Pathology Laboratory, Institute of Pathology, The Queen's University of Belfast, Grosvenor Road, Belfast BT12 6BL, N Ireland.

Accepted for publication 27 October 1993

Fine needle aspiration cytology (FNAC) has been established as a rapid, safe, and cost effective method of diagnosis in breast disease. The success of the technique relies strongly on the ability of the cytologist to identify and characterise cytological changes in the prepared aspirate. This presents a number of problems. Diagnosis is largely based on visual criteria which are subjective and can be misleading in certain instances. The number of visual clues which need to be assessed and the number of options available impose difficulties for assimilating all the relevant diagnostic information in a consistent and reproducible manner. This is also true of other areas of histological and cytological diagnosis.[1]

Expert systems are computer programs that are designed to store, access, and process knowledge about a particular domain.[2] They can therefore provide the perfect framework for storing cytological diagnostic knowledge in a logical, consistent, and reproducible manner and have substantial potential in providing cytopathologists with a means of making a more accurate and consistent diagnosis.

Decision making in cytopathology (as in other domains) involves the consideration and combination of numerous pieces of evidence. When the diagnostic evidence is clear-cut, the uncertainty involved in reaching a decision is low. However, most of the knowledge in diagnostic cytopathology is in the form of images, concepts, and descriptive terminology and because of this, uncertainty in the decisions leading to diagnosis and in the final diagnosis is inherent. It is for this very reason that experts are required to make difficult diagnostic decisions. The development of computer-based expert systems requires careful consideration as to how uncertainty is handled and how knowledge is to be represented within the design.[3][4]

Bayesian belief networks provide a powerful means of representing knowledge: evidence is combined in a cumulative manner to provide a measure of certainty in the final decision. Their structure is comprised of NODES connected by LINKS (fig 1A), where the nodes constitute probabilistic variables—nuclear pleomorphism—with possible outcomes—none, mild, moderate, severe—and links represent the relation between two nodes and are quantified by a conditional probability (CP) matrix. The CP matrix expresses the relation between the possible outcomes of the descendent child node with the outcomes of the parent node. When the descendent node represents evidence, such as a cytological feature, and the parent node is the diagnosis as in the current study, the CP matrix expresses the probability of finding a feature outcome (such as pleomorphism:

A

Conditional probability matrix   B

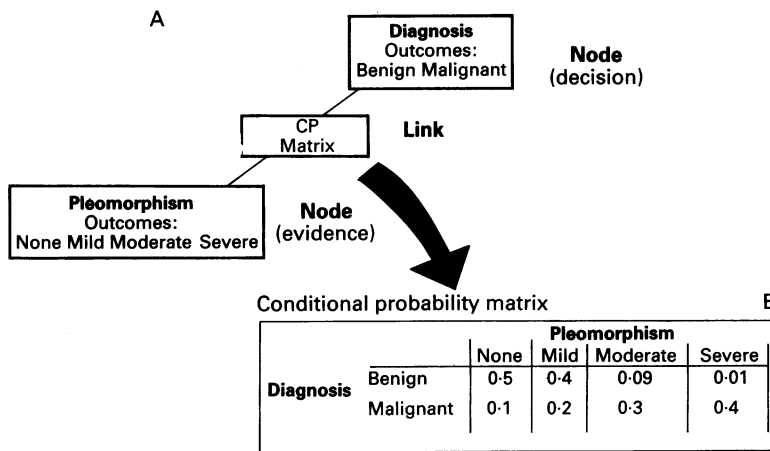|		**Pleomorphism**			
		None	Mild	Moderate	Severe
Diagnosis	Benign	0·5	0·4	0·09	0·01
	Malignant	0·1	0·2	0·3	0·4

Figure 1   (A) Diagram of basic Bayesian belief network unit comprising a decision node (DIAGNOSIS) linked to an evidence node (PLEOMORPHISM) via a conditional probability (CP) matrix. (B) This shows an enlargement of the CP matrix link which conveys the probability of pleomorphism outcomes being present given the diagnosis.

Table 1   List of diagnostic features and their possible outcomes used in construction of Bayesian belief network

Bare nuclei:
  numerous, moderate, few/absent
Cellularity:
  low, moderate, high
Cohesion:
  strong, moderate, poor
Pleomorphism:
  none, mild, moderate, severe
Cell arrangement:
  even/flat, irregular/overlap
Nuclear size:
  small, medium, large
Nucleoli:
  none, single, multiple
Intracytoplasmic lumina:
  present, absent
Aprocrine cells:
  present, absent
Mucinous background:
  present, absent

mild) given a diagnostic outcome (such as benign) (fig 1B).

The belief in the various outcomes is also stored at each node. This is expressed as a probability. For the decision node, this is defined by the user. For the evidence nodes, the outcome probabilities are defined by a network initialisation process and are largely dependent on the values held in the CP matrices.[4]

The system reported here is designed to be interactive in that the cytopathologist examines the cytological features of a case and enters evidence into the network in the form of relative likelihood ratios. These convey the likelihood that each of the possible feature outcomes are present in the sample under investigation (see table 5 for examples). Values of zero are not used as this presents computational problems. The value for each outcome is an independent estimate and the total does not have to equal 1·00. They therefore define in numerical form the subjective impression of the cytologist.

When evidence is submitted to an evidence node, the belief in the outcomes of that node are updated. This belief is then propagated, in our very simple network, upwards to the diagnostic decision node, where the belief in the diagnostic outcomes (benign and malignant) are updated. This update, in turn, triggers an
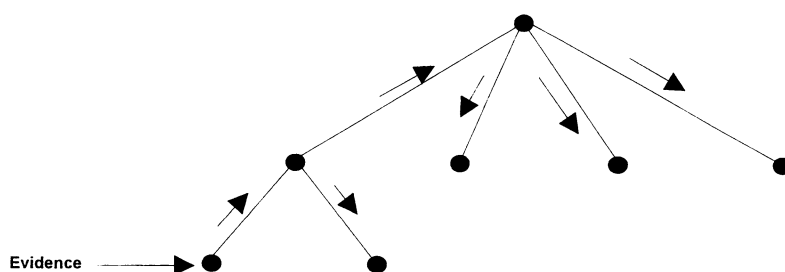
update of all the prior probabilities and beliefs at the other evidence nodes (fig 2). At the decision node, accumulating evidence will result in changes in the outcome belief until the belief in a particular outcome exceeds a defined threshold, sufficient to constitute a diagnostic decision.

A detailed description of the algorithmic mechanism of network initialisation and belief propagation in Bayesian belief networks is given by Bartels, Thompson, and Weber.[4] Bayesian belief networks have been applied in the diagnosis of prostate lesions[5 6] and their theoretical application in quantitative histopathology has been extensively researched by Bartels *et al.*[7]

Methods

BAYESIAN BELIEF NETWORK DESIGN
The Bayesian belief network was constructed using software developed in The Optical Sciences Centre, University of Arizona, USA. The program was written in C and has a graphical interface which permits the interactive definition of nodes, links, and CP matrices. Algorithmic implementation followed that designed by Morawski[8 9] from the concepts originally developed by Pearl.[10]

A set of 10 diagnostic clues were defined which provide evidence in the diagnosis of breast FNAC. These are listed in table 1 together with their possible outcomes. Note that these features were chosen for the assessment of aspirates stained using Hema-Gurr (BDH Chemicals), the method used in our laboratory where a rapid diagnostic service is provided at outpatient clinics.

The belief network topology was in the form of an open tree with a single level of evidence nodes (fig 3). This simple design consisted of a single decision node representing the DIAGNOSIS with its possible outcomes (BENIGN and MALIGNANT), and the 10 diagnostic features as evidence nodes directly linked to the diagnostic decision node. Conditional probability matrices relating each of the diagnostic clues and their outcomes to diagnosis (benign, malignant) were defined by a cytopathologist (NA) (table 2).

CASES
The performance of the belief network was tested using a series of 40 cases of breast aspirates retrieved from file at the Royal Victoria Hospital, Belfast, all of which had confirmatory tissue biopsy diagnoses. Based on the cytology report, these comprised 16 benign, 11 malignant, eight suspicious, and five atypical probably benign cases using the criteria of the NHS Breast Screening Programme[11] (table 3). Each case was assessed cytologically for all of the features listed in table 1 and relative likelihood ratios recorded and entered as evidence into each node of the network. After each piece of evidence was submitted the calculated probabilities for the diagnostic outcomes were recorded. After all the evidence nodes were sampled for a single case, the final diagnostic

Figure 2   Propagation of belief in a Bayesian belief network. ● = nodes, ——— = links. Arrows pointing up: update of likelihoods and beliefs. Arrows pointing down: update of prior probabilities and beliefs. This illustration shows a second level of descendent nodes at which the evidence was entered. The network in the current study only possessed a single layer of evidence nodes.

Figure 3 Topology of Bayesian belief network for the diagnosis of breast FNAC. The decision node is DIAGNOSIS and the evidence nodes are the cytological features. The outcomes for each node are also shown.
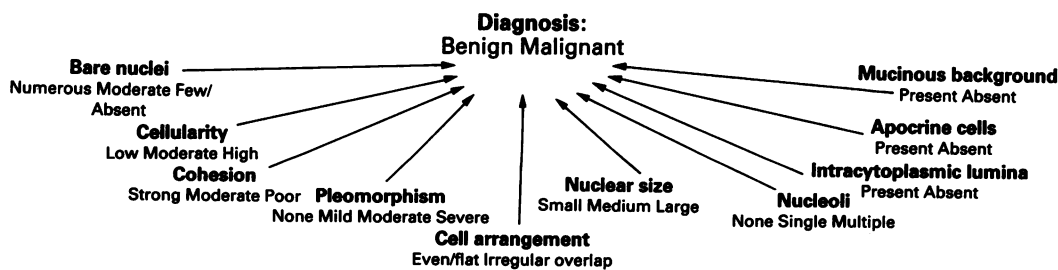
Diagnosis:
Benign Malignant

Bare nuclei
Numerous Moderate Few/
Absent

Mucinous background
Present Absent

Cellularity
Low Moderate High

Apocrine cells
Present Absent

Cohesion
Strong Moderate Poor **Pleomorphism**
None Mild Moderate Severe

Intracytoplasmic lumina
Present Absent

Nuclear size
Small Medium Large

Nucleoli
None Single Multiple

Cell arrangement
Even/flat Irregular overlap

probability was compared with the original diagnosis given by the cytopathologist.

NETWORK PERFORMANCE
The value of belief in the diagnostic outcomes, as provided by the network, was used to determine the ability of the system to diagnose correctly breast aspirates. Belief thresholds were defined so that cases could be allocated into groups based on their network derived diagnostic probability and compared with the original diagnosis. If a case had a probability of >0·90 it was recorded as a clear-cut BENIGN or a MALIGNANT case by the network. Cases which had a belief in benign of <0·90 but >0·50 were called EQUIVOCAL (B). Cases which had a belief in malignant of <0·90 but >0·50 were designated EQUIVOCAL (M).

ADDITIONAL DIAGNOSTIC MEASUREMENTS
Cumulative diagnostic probability graphs
Probability graphs were drawn for each case, plotting the cumulative diagnostic probability after each piece of evidence was submitted (fig 4). The evidence nodes were examined in an order considered by the cytologist to be of decreasing importance to the final diagnosis. Graph characteristics could be measured in simple terms: the number of peaks (P-score), troughs (T-score), and intersects with the 0·5, 0·5 threshold (C-score) were determined for each case. These measure the extent of conflicting information across the range of diagnostic clues.

Table 2 Conditional probability matrices relating probability of finding feature outcome, given the diagnosis (these sum to 1·0 in the horizontal)

	Bare nuclei		
	Numerous	Moderate	Few/absent
Benign	0·55	0·40	0·05
Malignant	0·01	0·10	0·89

	Cellularity		
	Low	Moderate	High
Benign	0·60	0·30	0·10
Malignant	0·05	0·10	0·85

	Cohesion		
	Strong	Moderate	Poor
Benign	0·70	0·25	0·05
Malignant	0·10	0·40	0·50

	Pleomorphism			
	None	Mild	Moderate	Severe
Benign	0·50	0·40	0·09	0·01
Malignant	0·10	0·20	0·30	0·40

	Cell arrangement	
	Even/flat	Irregular/overlap
Benign	0·90	0·10
Malignant	0·05	0·95

	Nuclear size		
	Small	Medium	Large
Benign	0·89	0·10	0·01
Malignant	0·20	0·40	0·40

	Nucleoli		
	None	Single	Multiple
Benign	0·74	0·25	0·01
Malignant	0·10	0·50	0·40

	Intracytoplasmic lumina	
	Present	Absent
Benign	0·01	0·99
Malignant	0·50	0·50

	Apocrine cells	
	Present	Absent
Benign	0·10	0·90
Malignant	0·01	0·99

	Mucinous background	
	Present	Absent
Benign	0·01	0·99
Malignant	0·05	0·95

Table 3 Diagnostic categories and summary of diagnostic criteria used in this study and defined by NHS Breast Screening Programme[11]

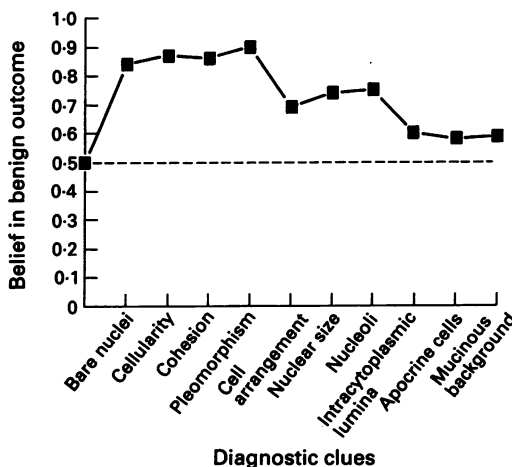| Benign	Indicates an adequate sample showing no evidence of malignancy
Atypical probably benign	The aspirate can have all the characteristics of a benign aspirate, but certain features not commonly seen in benign aspirates may be present
Suspicious of malignancy	The material is not diagnostic of malignancy but may show some malignant features without overt malignant cells being present
Malignant	Indicates an adequate sample containing cells characteristic of carcinoma or other malignancy

Figure 4 Cumulative probability graph for a benign fibroadenoma. Notice that it is the presence of bare nuclei which has the strongest initial effect in keeping the line in benign half of the chart.

Table 4 Binary code for absolute benign cases and for malignant cases

Bare nuclei			Cellularity			Cohesion			Pleomorphism				Cell arrangement	
Numerous	Moderate	Few/ Absent	Low	Moderate	High	Strong	Moderate	Poor	None	Mild	Moderate	Severe	Even/ flat	Irreg. overlap
						Absolute benign cases:								
1	0	0	1	0	0	1	0	0	1	0	0	0	1	0
						Malignant cases:								
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1

Binary coding and distance measures

This required defining a binary code for a hypothetical "absolute" benign case. For each diagnostic feature, the most typical outcome for a benign case was given a value of 1, and the other outcomes a value of 0 (table 4). Therefore, the binary code for benignity is: 100100100100101001000101 against which the codes for real cases were compared. Apocrine metaplasia was omitted from the binary scoring system, because while its presence is indicative of benignity, its absence has no diagnostic power. Each case in the series was allocated a similar code based on the most prevalent outcome of each feature (table 4).

Deviation away from the absolute normal binary code was quantified using two distance measures. Firstly, the Hamming distance measures the number of digits which change to produce the new code.[12] The greater the deviation away from the normal blueprint, the larger will be the Hamming distance.

A second measure was used where the sequences of diagnostic clue values were considered as a string, with substrings for each clue. The total number of jumps necessary to produce the new clue value for each clue independently were counted and summed. This we termed the "diagnostic clue distance". These measures were made for each case in the series as additional evidence for the diagnostic classification of cases.

Table 5 Likelihood ratios for a benign case number 1*

Bare nuclei:
  Numerous, moderate, few/absent
  0·30   0·90   0·30
Cellularity:
  Low, moderate, high
  0·75 0·65   0·10
Cohesion:
  Strong, moderate, poor
  0·90 0·60   0·05
Pleomorphism:
  None, mild, moderate, severe
  0·85 0·75 0·15   0·01
Cell arrangement:
  Even/flat, irregular/overlap
  0·80   0·10
Nuclear size:
  Small, medium, large
  0·95 0·25   0·01
Nucleoli:
  None, single, multiple
  0·90 0·50 0·10
Intracytoplasmic lumina:
  Present, absent
  0·10   0·95
Aprocrine cells:
  Present, absent
  0·01 0·99
Mucinous background:
  Present, absent
  0·10   0·98

*Entering this evidence into the Bayesian network resulted in a Benign/Malignant probability of 1·00/0·00 (see table 6)

REPRODUCIBILITY OF RELATIVE LIKELIHOOD RATIOS

The same series of cases were relabelled and presented to the same cytologist who was blind to the original diagnosis and to the previously defined likelihood ratios. They were reassessed and new likelihood ratios for each case entered into the network. Reproducibility of case classification based on the network diagnostic probability was assessed.

Results

Likelihood ratios for each of the diagnostic features were estimated for all 40 cases. An example of a typical result for a benign case is shown in table 5.

The diagnostic probability results provided by the Bayesian belief network for all 40 cases are shown in table 6.

Table 6 List of cases with resulting diagnostic probability provided by Bayesian belief network

	Belief	
	Benign	Malignant
Benign	1·00	0·00
	0·59	0·41
	1·00	0·00
	0·93	0·07
	1·00	0·00
	0·99	0·01
	0·99	0·01
	1·00	0·00
	1·00	0·00
	1·00	0·00
	0·99	0·01
	1·00	0·00
	1·00	0·00
	1·00	0·00
	1·00	0·00
	1·00	0·00
Malignant	0·00	1·00
	0·00	1·00
	0·08	0·92
	0·00	1·00
	0·00	1·00
	0·00	1·00
	0·00	1·00
	0·00	1·00
	0·00	1·00
	0·43	0·57
	0·00	1·00
Atypical/benign	1·00	0·00
	1·00	0·00
	0·98	0·02
	0·68	0·32
	0·95	0·05
Suspicious	0·91	0·09
	0·08	0·92
	0·12	0·88
	0·05	0·95
	0·84	0·16
	0·26	0·74
	0·00	1·00
	0·34	0·66

Nuclear size			Nucleoli			Intracytoplasmic lumina		Mucinous background	
Small	Medium	Large	None	Single	Multiple	Present	Absent	Present	Absent
				Absolute benign cases::					
1	0	0	1	0	0	0	1	0	1
				Malignant cases:					
0	1	0	0	1	0	1	0	0	1

BENIGN AND MALIGNANT CASES

Of the 16 cases diagnosed by the cytologist as benign, all were given probabilities in favour of a benign diagnosis by the network (table 6). One case, however, fell into the EQUIVO-CAL (B) category with a benign/malignant belief of 0·59/0·41. Review of this case showed that it was a fibroadenoma, a lesion which can pose diagnostic difficulties in cytology. Evidence for benignity came from the probability graph (fig 4) which showed that the probability line remained completely within the benign half of the graph. While the Hamming distance for this case had no discriminatory power, the diagnostic clue distance was lower in the fibroadenoma compared with the malignant cases (fig 5). None of the benign cases had a positive C-score (table 7).

Of the 11 malignant cases, again all were given probabilities in favour of a malignant diagnosis by the network (table 6). A single case had a lower probability than the rest (benign 0·43 malignant 0·57) putting it into the EQUIVOCAL (M) group. On review this was found to be an aspirate from an infiltrating lobular carcinoma, again a lesion difficult to diagnose from cytology. Examination of the probability curve (fig 6) clearly showed conflicting evidence with a C-value of 2. The number of evidence nodes which support a benign belief was four as opposed to six downward shifts supportive of malignancy.

This lobular carcinoma was the only malignant case to show a probability curve which cut the 0·50:0·50 threshold line (table 7).

ATYPICAL PROBABLY BENIGN CASES

All of the atypical probably benign cases had benign probabilities and one of these fell into the EQUIVOCAL (B) category. None of the belief curve characteristics or binary distance measures could be used to discriminate these from certain benign aspirates.

SUSPICIOUS CASES

Of the eight suspicious cases, four showed indeterminate probabilities meaning that they could not be directly classified as benign or malignant. Of these four, three showed higher probabilities for malignancy (EQUIVOCAL (M)). All of these three cases were confirmed to be malignant on biopsy. The single case having a network benign probability of 0·84 was shown to be an infiltrating lobular carcinoma. Three cases were classified as malignant by the network, all of which were confirmed to be malignant on biopsy. A single case was given a benign probability (>0·90) by the network, but examination of the probability curve showed it to have a C-score of 1, a T-score of 3 and a P-score of 3.

All but one of the suspicious cases had a probability curve which cut the 0·5:0·5 threshold on one or more occasions (table 7).

REPRODUCIBILITY OF RELATIVE LIKELIHOOD RATIOS

Re-examination of the same cases resulted in slightly different likelihood ratio vectors. Correct case classification based on the final probability, however, was not significantly affected. Benign and malignant cases were much more clearly defined on the second occasion, with all benign cases having a 1·00/0·00 benign/malignant probability and all malignant cases showing a 0·00/1·00 benign/malignant probability. As before,

Figure 5   Comparison of binary distance measures for benign and malignant cases. The fibroadenoma (F) which showed non-definite probability values was categorised into the benign range on the basis of the diagnostic clue distance.
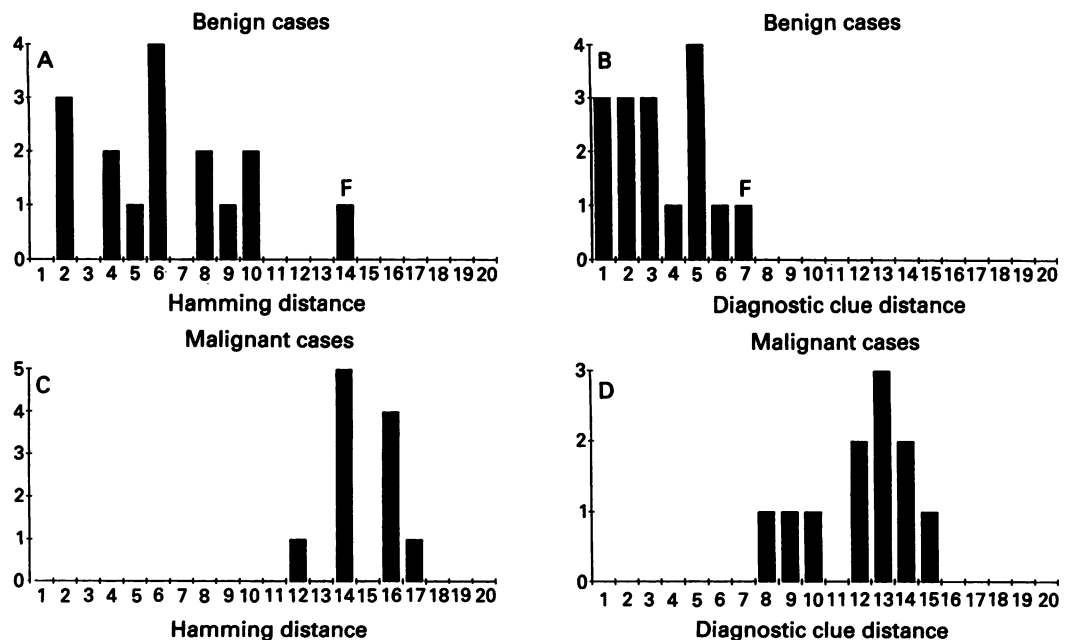
Table 7 Number of cases falling into C-score, P-score, and T-score intervals, as calculated from cumulative probability plots

	C-score			P-score					T-score				
	0	1	2	0	1	2	3	4	0	1	2	3	4
Benign (n = 16)	16	0	0	11	2	0	3	0	11	2	1	2	0
Atypical probably benign (n = 5)	5	0	0	3	0	0	2	0	3	0	1	1	0
Suspicious (n = 8)	1	3	4	0	1	4	2	1	0	1	1	5	1
Malignant (n = 11)	10	1	0	8	1	2	0	0	8	1	0	2	0

Figure 6 Cumulative probability graph for malignant lobular carcinoma.

Diagnostic clues
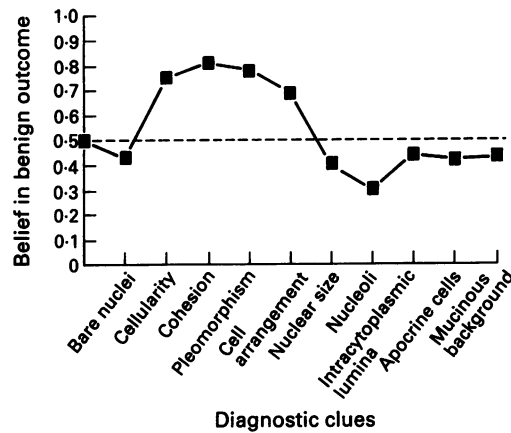
suspicious cases fell into the equivocal category with clear signs of conflicting evidence in their cumulative probability curves.

Discussion

The role of expert systems in the diagnosis of breast pathology has been considered before. Heathfield et al developed a rule based expert system for the diagnosis of breast FNAC based on a series of IF . . . THEN rules where uncertainty in individual rules and in the final decision was not assessed.[13] This approach is a commonly used method of representing knowledge in expert system development, but it is questionable whether diagnostic expertise can be reduced to a series of simple monotonic rules. The incorporation of certainty measures in rule based systems, the methods used to update certainty and the handling of missing data, can also be problematic.[10 14] Bayesian belief networks, however, can provide a reliable method for handling certainty which is based on well established probability calculus. They can also operate to completion and reach a final diagnostic belief even when evidence is missing. Some of the problems associated with rule based knowledge representation were recognised in a later paper by Heathfield's group[15] who had developed an expert system for diagnosis in breast histopathology.

The relation between cytological features and the diagnosis is clearly associated with uncertainty. In Bayesian models this uncertainty is represented by conditional probabilities which express the likelihood of a feature being present given the diagnosis. In theory, it might be possible to use statistically defined probabilities—from frequency counts—to construct the CP matrices necessary for Bayesian belief networks. In the current

study, however, CP matrices were constructed using probabilities proposed by the cytopathologist—that is, "personal probabilities". They are an attempt to convey in quantitative terms the knowledge and experience of the cytopathologist and as such, provide useful information to a knowledge based system of this kind. CP matrices are central to the success of a belief network and care must be taken in their construction. If the defined conditional probabilities do not provide the expected results they can be adjusted in a training process, based on specimens of known diagnosis, until the network performs correctly. In this particular study, however, the original CP matrices devised by the cytopathologist remained unchanged even after testing.

To start the network, prior probabilities must be given for the diagnostic outcomes, benign and malignant. In practice, the attitude adopted by most cytopathologists is that each aspirate is examined equally and independently, with no preconceived notions as to its most probable diagnosis, regardless of the fact that in statistical terms an individual case is more likely to be benign than malignant. We therefore used equal prior probabilities (0·5,0·5) for the diagnostic outcomes of benign and malignant in the current network and this was shown to work well.

It was evident that while the results of the final belief scores were useful, additional information could be obtained by plotting the cumulative diagnostic probability curve after each evidence node had been sampled. This allowed the cytopathologist to trace the steps made in the diagnostic decision sequence and their effect on the final decision, a useful procedure, particularly for difficult cases. It also highlighted evidence which did not match the general trend of the curve. If doubt existed over the response of the curve to the evidence entered for a particular feature then this feature could be reviewed cytologically. This quantitative feedback on the diagnostic process is fundamental to ensuring that each piece of diagnostic evidence is considered in full and that diagnostic accuracy is improved. The storage of diagnostic probability curves in this way provides a quantitative record of the diagnostic decision process and would have important implications in assessing reproducibility and quality control. It is important to remember that the quantitative characteristics of the probability curve are strongly dependent on the order in which the evidence nodes are sampled. This needs to be standardised, as in this study, if comparisons are

to be made. It should be emphasised, however, that the final probability is completely independent of the order in which the evidence nodes are examined.

The ability of the network to allocate correctly benign and malignant breast aspirates on the basis of the diagnostic probability was excellent and when apparent misclassifications occurred, it was due to lesions which are known to pose diagnostic difficulties. Only one case of fibroadenoma (out of four) fell within the equivocal belief range. Supportive evidence of the benign nature of this case, however, came from the probability curve characteristics. The fact that bare nuclei are such an important discriminant in the differential diagnosis of fibroadenoma and carcinoma perhaps justifies assigning a stronger CP matrix for this feature in future. The diagnostic clue binary distance measurement also showed this case to be more closely akin to the other benign cases than malignancy.

The single malignant case which was classified into the equivocal range by the network was an aspirate from an infiltrating lobular carcinoma. Examination of the cumulative probability curve for this case (fig 7) also clearly showed the presence of conflicting evidence. It is encouraging that the network was capable of firstly allocating this case in the equivocal (M) belief range and secondly, the high C-score would alert the user to the unusual nature of the evidence.

Atypical probably benign cases all fell into the benign belief range and could not be differentiated from true benign cases. In practice this could be satisfactory as atypia is often seen in examples of benign epithelial hyperplasia. All of the atypical cases were benign on biopsy.

Suspicious cases could be characterised not only by their equivocal probabilities but by their probability curves which were highly erratic, resulting in high C, T, and P values. Success of the network diagnostic probability was also evident as suspicious cases which were classified as EQUIVOCAL (M) or MALIGNANT by the network were confirmed to be malignant on biopsy.

The reproducibility of individual likelihood ratios for each feature was questionable, but the final classification of cases in the repeated study was highly reproducible. This demonstrates the need for multiple evidence in reaching a diagnostic decision and shows that no single feature provides decisive evidence. Bayesian belief networks permit the combination of features which independently may be weak but cumulatively lead to a decision with high certainty. While people are capable of combining evidence to support a decision, this is often undermined by conceptual, terminological, and behavioural inconsistencies and they cannot quantify the process leading to, and the certainty of, the final decision. Bayesian belief networks, while maintaining the role for linguistic and descriptive terminology for evidence communication, provide a quantitative means of describing the diagnostic decision making process.

These results verify the potential of this system as an aid to decision making. The network needs to be tested on a larger series of cases with emphasis on those cases which pose particular difficulties—for example, lobular carcinomas. Only by examining larger numbers of difficult cases can methods to improve network performance be found.

In the current design only cytological evidence has been used. In practice, clinical evidence provides important supplementary data used in the diagnosis. Future work will incorporate clinical evidence either within the existing network structure or as a separate network whose result can be used to support or reject the outcome of the cytological belief network. Also the network described here is based on features seen in air-dried slides but a similar network could be constructed using features appropriate for alcohol fixed preparations.

The value of the system designed in the current study is three-fold. Firstly, it can be used to provide the cytopathologist with an aid to decision making in the routine diagnostic assessment of breast aspirates. At the moment the design of the system is primitive in that the cytopathologist has to provide numbers (likelihood ratios) for each evidence node. This can be time consuming and even with practice can still take between five and 10 minutes per case. This could be overcome by using stored image libraries with predefined likelihood ratios, allowing the user to select images which most closely match the features seen in the current case. This divorces the user from having to convert his or her cytological impressions into numbers yet provides a highly visual means of providing data suitable for the network. As cumulative probability curves for cases with certain outcomes are recorded, statistical data will become available on conditional probability values between P, T, and C scores and the diagnostic outcome. The network could be extended to use this information about its own decision making process as additional diagnostic clues, and enter likelihood vectors as a last step before it reports a final outcome probability.

While an interactive system as described might prove successful, the role of quantitative cytomorphometric data should not be negated. Numerous studies have shown that nuclear and cellular measurement can be beneficial in the objective diagnosis of breast cytology[16-20] and future work should attempt to include these as additional evidence within the framework of a Bayesian belief network.

Secondly, a model of this sort can be used as the basis for a teaching system for those inexperienced in breast FNAC diagnosis. The computer could, therefore, act as a consultant, leading the user through the diagnostic process, prompting for information about the case in hand, providing visual examples of diagnostic features, querying the users' responses and on reaching a decision, retrace the steps used to obtain it.

Finally, it would be desirable to design a

computer system capable of automated diagnosis in breast FNAC. Such systems have been considered for many years in cervical cytology[21-23] and recent increases in processing power and advances in machine vision, expert system, and neural network technology have now made this a reality.

It is clear that expert computer systems are going to be increasingly investigated in diagnostic histology and cytology in the next few years. Computerised interactive decision support systems, such as the one described here, will, without doubt, lead to improved consistency and reproducibility in diagnosis.

1 Smeulders AWM, van Ginneken AM. An analysis of pathology knowledge and decision making for the development of artificial intelligence-based consulting systems. *Anal Quant Cytol Histol* 1989;11:154-65.
2 Bartels PH, Hiessl H. Expert systems in histopathology. II Knowledge representation and rule based systems. *Anal Quant Cytol Histol* 1989;11:147-53.
3 van Ginneken AM, Smeulders AWM. Reasoning in uncertainties: An analysis of five strategies and their suitability in pathology. *Anal Quant Cytol Histol* 1991;13:93-109.
4 Bartels PH, Thompson D, Weber JE. Expert systems in histopathology. IV The management of uncertainty. *Anal Quant Cytol Histol* 1992;14:1-13.
5 Bibbo M, Bartels PH, Pfeifer T, Thompson D, Minimo C, Galera Davidson H. Belief network for grading prostate lesions. *Anal Quant Cytol Histol* 1993;15:124-35.
6 Montironi R, Bartels PH, Thompson D, Scarpelli M, Hamilton PW. Prostate intraepithelial neoplasia: development of Bayesian belief network for diagnosis and grading. *Anal Quant Cytol Histol* (in press).
7 Bartels PH, Thompson D, Bibbo M, Weber JE. Bayesian belief networks in quantitative histopathology. *Anal Quant Cytol Histol* 1992;14:459-73.
8 Morawski P. Understanding Bayesian belief networks. *AI Expert* 1989;**August**:44-8.
9 Morawski P. Programming Bayesian belief networks. *AI Expert* 1989;**May**:74-9.
10 Pearl J. *Probabilistic reasoning in intelligent systems.* California, Morgan Kaufman, 1988.
11 DoH. NHS Breast Screening Programme. *Guidelines for Cytology Procedures and Reporting in Breast Cancer Screening. NHSBSP Publication No 22.* London: HMSO, 1992.
12 MacWilliams FJ, Sloane NJA. *The theory of error-correcting codes. Part I.* Amsterdam: North Holland Publishing Company, 1977.
13 Heathfield HA, Kirkham N, Ellis IO, Winstanley G. Computer assisted diagnosis of fine needle aspirate of the breast. *J Clin Pathol* 1990;43:168-70.
14 Bartels PH, Thompson D, Weber JE. Diagnostic decision support by inference networks. *In Vivo* 1993;7:379-86.
15 Heathfield H, Bose D, Kirkham N. Knowledge-based computer system to aid in the histopathological diagnosis of breast disease. *J Clin Pathol* 1991;44:502-8.
16 Cornelisse CJ, de Loning HR, Arentz PW, Ratgever JW, van Heerde P. Quantitative analysis of the nuclear area variation in benign and malignant breast cytology. *Anal Quant Cytol* 1981;3:128-34.
17 Boon ME, Trott PA, van Kamm H, Kurver PJH, Leach A, Baak JPA. Morphometry and cytodiagnosis of breast lesions. *Virchows Arch (Pathol Anat)* 1982;**396**:9-18.
18 Wolberg WH, Tanner MA, Loh W-Y, Vanichsetakul N. Statistical approach to fine needle aspiration diagnosis of breast masses. *Acta Cytol* 1987;31:737-41.
19 van Diest PJ, Risse EK, Schipper NW, Baak JP, Mariquand J. Comparison of light microscopic grading and morphometric features in cytological breast cancer specimens. *Path Res Pract* 1989;**185**:612-16.
20 Beerman H, Veldhuizen RW, Blok RA, Hermans J, Ooms EC. Cytomorphometry as quality control for fine needle aspiration. A study in 321 breast lesions. *Anal Quant Cytol Histol* 1991;13:143-8.
21 Tang Z, Savino A, Wong EK, Koss LG, Shaw LG. An expert system designed as a tutoring tool in cervical cytology: TTCC-1 system. *Anal Quant Cytol Histol* 1988;10:417-22.
22 Hall T. Detection of abnormal cells in cervical smears using a neural network-based pattern recognition system. *Anal Quant Cytol Histol* 1990;12:194-5.
23 Anonymous. Data on automated cytology systems as submitted by their developers. [Editorial]. *Anal Quant Cytol Histol* 1991;13:300-6.