Viewpoint ■

Are Normative Expert Systems Appropriate for Diagnostic Pathology?

LAWRENCE W. DIAMOND, MD, VLADIMIR G. MISHKA, PhD, APOLLO H. SEAL, PhD,
DOYEN T. NGUYEN, MD

Abstract Conceptual models for diagnostic reasoning proposed in the medical literature are presented to stimulate discussion about the issue of the appropriateness of probabilistic knowledge-based systems for medical diagnosis. Evidence is presented to corroborate the authors' view that diagnosis is a problem-solving task, rather than a decision-making task. In the authors' opinion, probabilistic reasoning is better suited to situations dealing with choices for clinical intervention, rather than to those dealing with determining the correct diagnosis. A critique is given of a diagnostic Bayesian expert system for lymph node pathology. In empirical studies, diagnostic Bayesian systems have been shown to typically list the correct diagnosis as the program's first choice 60% to 70% of the time. One reason for this undistinguished level of diagnostic performance is that Bayesian systems are not designed to represent and use knowledge the same way that an expert does.

■ *J Am Med Informatics Assoc.* 1995;2:85-93.

Knowledge-based computer systems for medical diagnosis have been constructed since the early 1960s,¹ but their promise remains largely unfulfilled.² The failure to integrate knowledge-based systems (KBS) into medical practice has been identified as one of the main reasons for the lack of demonstrable success of this technology in health care.³ Many developers are trying to widen the scope of KBS into more general patient management, since diagnosis plays only a partial role in medicine.

There are, however, circumstances in which diagnosis is vitally important. The early diagnosis of cancer in biopsy specimens and the proper diagnosis and subclassification of leukemias and malignant lymphomas greatly influence the choice of appropriate therapy and can make the difference between a potential cure and the death of the patient. Why have KBS not realized their potential in purely di-

agnostic tasks, and why are they not in widespread use in diagnostic pathology?

It has often been said that "medical knowledge and reasoning are replete with uncertainty" and that medical decision support systems must be capable of representing and reasoning with uncertainty.⁴ This has led to a great deal of research on probabilistic and decision-theoretic systems utilizing belief networks, or decision networks with utility variables and utility functions. The clinician is seen as a decision maker, and clinical practice is viewed as a series of lotteries or gambles. To quote Elstein, "The choices of the decision maker, whether among diagnostic tests or between treatment alternatives, are thought of as decisions to participate in one lottery or another. The metaphor [of medical decision making] emphasizes the uncertainties of clinical interventions rather than the need to determine the causes of the current situation."⁵

These same ideas about medical decision making are being applied to the field of diagnostic pathology. Considerable effort has been addressed to the development of normative expert systems for fields such as interpretation of peripheral blood findings and lymph node pathology.^{6,7} In these systems, knowledge is represented in the form of thousands of subjective conditional probabilities.

Affiliation of the authors: Pathology Institute, University of Cologne, Cologne, Germany.

Correspondence and reprints: Lawrence W. Diamond, MD, Department of Haematology, St. Bartholomew's Hospital, West Smithfield, London EC1A 7BE, United Kingdom.

Received for publication: 9/21/94; accepted for publication: 11/16/94.

Are diagnostic and treatment decisions made in the same manner, or are diagnostic decisions in pathology different from those in other domains of medicine? Should diagnosis in pathology be modeled as a lottery? Do subjective conditional probabilities effectively capture the domain knowledge in pathology? Only by openly debating these issues and critically examining the normative expert systems that have been developed for diagnosis can we learn when medical knowledge should be modeled according to probability theory and when other methods of knowledge representation are more appropriate.⁸ In this article, we take the position that diagnosis, especially in pathology, is different from clinical decision making. From the experience we have gained, both in the practice of hematopathology and in the development of a hematology workstation, we believe (and our experience supports) that real-life diagnoses are made by reasoning like a detective, not like a gambler.

How Is Diagnosis Different from Decision Making?

Deber and Baumann have presented an insightful examination of clinical reasoning in medicine, pointing out the distinction between problem solving and decision making.⁹ Using their terminology, problem solving is the search for the one "correct" answer and decision making refers to a situation in which a choice must be made from several possible alternatives. Decision making often requires trade-offs where utility information is valuable.

Determining the correct diagnosis is problem solving. Medical problem-solving tasks tend to be retrospective. They have been likened to a coin that has already been flipped or a horse race that has already been run.^{10,11} It is the diagnostician's job to uncover the evidence for a disease process that has already taken place and to classify the disease, i.e., assign a diagnostic name that provides a logical explanation of the findings. Diagnosis is not about clinical intervention, remembering Elstein's words quoted above, but rather it is about determining the cause of the current situation. Treatment selection, on the other hand, is an example of medical decision making. It is prospective, involving a choice about what to do next. When choosing a treatment modality, one needs to identify the alternatives and the probability of each likely outcome and to assign utility values for each of the outcomes. As has been pointed out, "it is only when faced with decision-making tasks that it becomes necessary to ascertain patient values and preferences."⁹

Szolovits and Pauker suggest that any program that will ultimately demonstrate expertise in as broad a field as internal medicine will have to use a judicious combination of both categorical and probabilistic reasoning.¹² As they indicate, it is essential to use each method at the appropriate time. They correctly point out that probabilistic methods are appropriate for addressing problems such as which treatment should be selected and how much of a drug should be given. They state that it is the "role of categorical methods to discover what the central problem is and to limit it as strongly as possible; only then are probabilistic techniques appropriate for its solution."¹² To the pathologist, diagnosis is the problem-solving task by which he or she identifies what the problem is and limits it whenever possible by classifying it. Szolovits and Pauker conclude their discussion of the role of categorical and probabilistic reasoning in medical diagnosis by emphasizing that "although probabilistic techniques will be best in some well-defined domains, they should not be applied arbitrarily to making other decisions where the development of precise categorical models could lead to significantly better performance."¹²

What Are the Fundamental Processes of Medical Diagnosis?

Deber and Baumann have also used empirical evidence, based on three separate studies, to support the view that problem solving, in clinical medicine and nursing, is based primarily on pattern recognition following data gathering.⁹ They suggest that a pattern recognition approach, stressing the importance of a knowledge base, "implies a strong potential role for the computer in both medical practice and medical education."⁹ Bartels has also pointed out that patterns are a fundamental concept in human diagnostic capability.¹³ Patterns arise at many levels and are not restricted to visual patterns that a pathologist sees under the microscope. Patterns are present in the diagnostic knowledge of expert physicians relating image data, patient clinical findings, laboratory test data, and knowledge of disease processes.¹³ Our own empirical evidence points to the same conclusion. While working on educational computer applications for the hematology laboratory, we set out to determine how one could improve on the diagnostic accuracy of Professor Claude Sultan's existing Bayesian system⁶ for hemogram and peripheral blood smear interpretation. To simplify the experiment, we used Bayes' rule. We tested three systems for hemogram analysis¹⁴: 1) the existing system that attempted to reach a final diagnosis⁶; 2) a system attempting to reach a less specific diagnosis based

on textbook categories of disease; and 3) a system attempting to reason on the basis of defined diagnostic patterns. The system based on defined patterns was clearly superior to the others. After working with Bayes' rule, we realized that our pattern approach would work better as a heuristic classification system as defined by Clancey.¹⁵ We have maintained this philosophy (heuristic classification using defined diagnostic patterns), not only in the development of a new system for hemogram and peripheral blood smear interpretation,^{16,17} but also in the development of the other modules in our hematology workstation.¹⁸ We have added modules for flow cytometry immunophenotyping and DNA content analysis,^{19,20} bone marrow pathology,^{21,22} and lymph node interpretation²³ to the workstation. The four modules communicate by means of a relational database,²⁴ which includes clinical information relevant to hematology diagnosis.²⁵

Feinstein has provided an in-depth analysis of diagnostic reasoning in medicine in a series of papers published more than 20 years ago.²⁶⁻²⁸ In that series, he demonstrated that "diagnosis is a rational activity of clinical medicine, rather than a computational exercise in numerical statistics."²⁶ Every clinician knows that the "end point" of diagnosis may vary depending on the clinical situation. A clinician "may be interested not in finding a precise diagnostic identification, but in ruling out the possible presence of a disease that can be cured or ameliorated."²⁶ Diagnostic end points are moving targets in surgical pathology, as well. If the current therapy for several different morphologic categories of aggressive non-Hodgkin's lymphoma is the same, then it may be acceptable to make a diagnosis simply of "high-grade, B-cell, non-Hodgkin's lymphoma" rather than to distinguish between the large cell and immunoblastic subtypes. We believe that the ability to appropriately stop at the more general diagnosis is an important attribute of a diagnostic system. Bayesian systems are often designed to concentrate the user's attention on the probability of specific diagnoses.^{6,29} It may be more difficult in Bayesian systems than in other approaches to include both general diagnoses and specific diagnoses at the same time, since the two types of diagnoses overlap. Including overlapping diagnoses in a Bayesian system may lead to a violation of the "requirement" for conditional independence.

Feinstein has also pointed out the importance of a strategy of intermediate decisions in diagnostic reasoning.²⁷ Despite his use of the term "decisions," which cannot be avoided in a discussion of diagnosis, he was clearly writing about problem solving. These intermediate decisions, or "intellectual stations" as

he called them, correspond to the patterns that are the fundamental building blocks of our approach to KBS design. Diagnosis should be decomposed into a series of steps. For each step, the physician should try to attain complete certainty. If an intellectual station cannot be reached with certainty, a plan of action is formulated, usually to obtain additional data.²⁷

Are "Clinical Diagnosis" Systems Examples of Problem Solving or Decision Making?

The system of de Dombal et al. is often quoted as an example of a successful Bayesian system for clinical diagnosis.³⁰ Using conditional probabilities previously collated from actual patients, the computer system's clinical diagnosis matched the final pathologic diagnosis in 279 of 304 cases (91.8%). This performance was better than that of the most senior member of the clinical team who saw each case.³⁰ In contrast, when a clinician's estimates of conditional probabilities were used instead of "real-life" data, the "system was often less effective than the unaided clinician—especially when diagnosing less common disorders."³¹

We agree that this level of performance is impressive, but we view the clinical diagnosis of abdominal pain as a decision-making task, quite different from the problem-solving task of pathologic diagnosis. The clinical evaluation of acute abdominal pain involves an immediate treatment decision, i.e., should the patient be subjected to surgery (either as a diagnostic test or for therapy), or is the best management non-surgical intervention? In patients with acute abdominal pain who undergo surgery, the actual diagnosis is made not by the clinician, but by the pathologist. In our opinion, the program for abdominal pain should be viewed as a system for aiding in treatment decisions. As noted above, probabilistic systems may be quite appropriate for decision-making tasks.

Can We Overcome the Need for Representations of Uncertainty in Diagnostic Pathology Systems?

Representations of uncertainty in computer-assisted medical diagnosis research have been around for a long time. Most investigators have felt the need for a mechanism to deal with uncertainty because of problems stemming from either 1) the incomplete nature of the true patient description, or 2) the inherent chance nature of disease processes.³² A number of schemes have been proposed and tested, including 1) Bayes' theorem,^{6-8,29-31} 2) decision analysis

with utility estimates,³³ 3) MYCIN certainty factors,³⁴ 4) Dempster-Shafer theory,³⁵ and 5) fuzzy set logic.³⁶

We do not dispute a role for representations of uncertainty in prospective decisions, involving a choice as to what diagnostic procedure to do next or what treatment option to choose. Probabilistic reasoning definitely has a place in medical decision making, but even so, it has its limitations. Individuals well versed in statistics remind us that probabilistic approaches apply to populations of patients, but may not be appropriate for individual patients.²⁷ Everyone is familiar with the situation where a cardiac surgeon informs a patient, the night before open heart surgery, that about 10% of patients who have undergone this type of operation have died on the operating table. It is unclear what the information means to that particular patient or his or her family, however. For an individual patient, there are two possible outcomes: life or death.

Some would argue that any method of knowledge representation that uses general knowledge about a branch of medicine to make inferences relevant to a particular patient will risk not being applicable to individual patients. We propose that our pattern-based approach does not suffer from this problem. From our review of the literature, many KBS developers have viewed medical syndromes and diseases as being potentially characterized by hundreds, if not thousands, of possible signs, symptoms, and laboratory test values. Many developers appear to be concerned with what features can be associated with a disease and not with what set of features must be present to define an entity. In our approach, each pattern is strictly defined. When definitions are used, no matter how arbitrary the definitions may seem to any particular observer, the system is guaranteed to be valid across all patients within the framework of those definitions. Furthermore, the need for representing uncertainty is eliminated.

The expert systems developed within a framework of criterion tables³⁷ employ heuristic reasoning in a manner similar to our defined diagnostic patterns. Indeed, some of the earliest work on knowledge-based approaches using defined criteria was in our own field, hematology. In the late 1950s, Lipkin and Hardy³⁸ showed that a mechanical system using marginal punched cards could achieve a diagnostic accuracy rate of greater than 90%, and could offer the ability to identify any important information needed to confirm the diagnosis. Later, Lipkin et al. developed a computer program for differential diagnosis in hematology that was one of the earliest systems to employ an approach similar to the one that we have adopted.¹

The certainty factor (CF) model was developed by Shortliffe and Buchanan for use in MYCIN, a rule-based system for the diagnosis and treatment of meningitis and bacteremia.³⁴ This model became part of the EMYCIN expert system shell and has been extensively used by the medical informatics community.

Whatever the role of numerical CFs in decision making, there is empirical evidence to suggest that they do not necessarily improve the performance of some problem-solving systems. Alvey et al. set out to develop a high-performance expert system capable of diagnosing leukemia subtypes from monoclonal antibody test results.³⁹ Their system started out as an EMYCIN system that was later translated directly into PROLOG. At one point in the development of their system, they decided to test whether the EMYCIN-style uncertainty mechanism contributed anything to the correctly diagnosed cases. A copy of the PROLOG system was made with the numerical uncertainty values removed. When the modified system was run, it produced exactly the same results as before. Without a mechanism for dealing with uncertainty, the system was able to give an acceptable conclusion on each of 400 cases tested, along with summaries of its reasoning and suggestions for further tests.⁴⁰

We have concentrated on categorical knowledge representation by breaking diagnostic problems down into more manageable steps. In our KBS from the diagnostic hematology workstation, there are no representations of uncertainty. Every intermediate conclusion is based on defined diagnostic patterns, and a complete interpretive report is generated using "if. . . , then" rules, which fire only within a given pattern when specific findings are present. The report takes into consideration the available clinical information and the hematology tests that have already been conducted on the patient. For example, our bone marrow module, known as Professor Belmonte,^{21,22} in addition to considering the physician's data from the examination of bone marrow smears and biopsy sections, automatically searches a relational database²⁴ for the results of a hemogram, peripheral blood smear findings, and flow cytometry immunophenotyping results on blood or bone marrow. The system correlates any information retrieved from the database with the bone marrow findings to generate a multiparameter interpretive report.^{22,41} We recently evaluated the accuracy of the bone marrow system's interpretive report and diagnosis for 100 consecutive patients with leukemia. The system's diagnosis agreed with the physician's diagnosis in 99 of 100 cases.⁴¹

Do Normative Pathology Expert Systems Reflect the Diagnostic Reasoning of the Domain Expert?

The Pathfinder project and the Bayesian expert system for peripheral blood analysis designed by the late Professor Claude Sultan both started life as rule-based systems.^{42,43} Intuitively, the rule-based logic approach was seen as the preferred method of acquiring knowledge. Predictably, the accuracy of the diagnoses based on the first attempts at writing a few hundred rules and entering them into an expert system shell was dismal. Both of these diagnostic tasks are far too complex for rules that all exist in the same pool and can "fire" at the same level. To successfully acquire sufficient knowledge to model hematopathology using a rule-based or similar approach, one needs to gain a deep understanding of the patterns that exist and how an expert uses those patterns to arrive at a diagnosis.

The developers of Pathfinder have argued that the surgical pathologist's role is to sequentially identify relevant microscopic features that appear in tissue sections.⁷ They have modeled the task as one in which the goal is to decide which features to look for until a diagnosis is reached. Looking for the next feature is viewed as if it were equivalent to performing another diagnostic test, thus requiring a decision about what to do next. In their model, there is uncertainty as to which feature to look for next, or which special stain or special procedure to perform. According to the definitions we have quoted above, their model reflects the idea that diagnostic lymph node pathology is a decision-making task, not a problem-solving task. Therefore, they use decision theory and the rules of probability for manipulating and managing uncertainty.

Does that view of lymph node pathology reflect the strategy actually used and taught by the domain expert? The methods for morphologic lymph node interpretation advocated by the Pathfinder domain expert are well documented in the literature.⁴⁴⁻⁴⁶ The proper approach to morphologic lymph node diagnosis is based on the search for patterns at low-power examination on well-prepared lymph node sections. Such patterns correspond to specific sets of differential diagnoses. A number of patterns have been established that can be recognized at low-power magnification, by modulating the light intensity and carefully observing the colors and architecture within the sections. The Pathfinder expert's knowledge of lymph node morphology is largely contained in his carefully defined histologic patterns.

The patterns are certainly discussed in the didactic portion of Intellipath, the commercial version of Pathfinder.⁴⁴ But where are they in the normative expert system itself? They are mostly lost, buried among 146 "features" that can be entered in random order, if desired. At the user's request, the system will recommend which crucial features to look for, using one of three different strategies: 1) to distinguish among groups of diseases (i.e., benign lymphadenopathies, hematopoietic malignancies, and other diseases such as AIDS or metastatic carcinoma); 2) to distinguish among all diseases ungrouped; and 3) to pursue the leading disease (i.e., the disease with the highest probability at the moment). A pathologist using the system can chase after the "leading disease" feature by feature, changing tack as the probabilities for certain diseases rise and fall. However, the system never directs the user to first establish the pattern of the lymph node as the domain expert would.

Why might Pathfinder not closely reflect the expert's knowledge representation scheme of low-power histologic patterns?^{44,45} It is a well-known principle of expert system design that "the knowledge engineer's knowledge, rather than the expert's is actually reflected in an expert system."⁴³ The knowledge engineer must have a deep understanding of the domain and must develop a conceptual model for it. Since the knowledge engineer's model actually underlies the construction of the formal knowledge base, problems can occur if that model differs from the model of the domain expert.

What Is the Appropriate "Gold Standard" for Evaluating Diagnostic Pathology Systems?

The choice of a "gold standard" is a problem that developers struggle with when evaluating KBS. Without an appropriate "gold standard," it is difficult to judge the efficacy of systems in real-life situations.

It has been claimed that "in the pathology domains, there appears to be no ideal gold standard."⁴⁹ The obvious "gold standard," the correct diagnosis, has been asserted to be "unworkable" for lymph node diagnosis because "nonexperts" have difficulties in recognizing features in the lymph node domain, i.e., they "misrecognize or fail to recognize some of the features associated with diagnosis."⁴⁹ Misinterpretation of morphologic features is a known problem in lymph node diagnosis, particularly when the lymph node sections being studied are of suboptimal quality.⁴⁶ We believe, however, that the correct diagnosis (or at least the diagnosis of record, if the case has been diagnosed by an expert) is a very appropriate

standard on which to evaluate the accuracy of any diagnostic system.

How Well Do Normative Pathology Systems Perform When Compared with the Diagnosis of Record?

In evaluating the diagnostic accuracy of the Pathfinder system, two versions of Pathfinder were used; one based on a simple Bayes model and one based on a belief network that represents dependencies among observations.⁴⁹ The diagnostic accuracy of Pathfinder was stated to be "at least as good as that of the Pathfinder expert" based on an evaluation of 53 cases.⁴⁹ The diagnosis of record was not used as the "gold standard" in making that claim, however. The domain expert was shown the values input into the two systems. Then, he was asked to rate the output of each system on a scale of one to ten. This subjective rating was one of the metrics used to judge the diagnostic ability of the systems.

Since two of us (DTN, LWD) were involved in the study evaluating the diagnostic accuracy of Pathfinder,⁴⁹ the data, including the diagnosis of record in all 53 cases, have been kindly made available to us. We have checked how often the Pathfinder system's first-listed choice matched the diagnosis of record. The simpler version matched the original diagnosis in 32 cases (60.4%) and the complex version matched the diagnosis in 36 cases (67.9%). These values are comparable with those of other Bayesian systems designed for diagnosis.^{6,34} When Professor Sultan's system for evaluating complete blood counts was evaluated with 180 cases of anemia, the correct diagnosis was displayed in the first rank in 64.5% of the cases.⁶

Recently, when a probabilistic system for the diagnosis of renal masses was evaluated using three different inference strategies, the three strategies had accuracies ranging from 61% to 64% when highly specific tests such as angiograms, computerized tomography, and magnetic resonance imaging were not considered and accuracies ranging from 69% to 76% when all of the information was available to the system.⁵⁰ The authors of the renal mass study also looked at the reliability and discriminating power of the diagnoses reached by their systems. They concluded that for the probabilistic models that they studied, "perfect performance in all of these parameters may not be easily attainable."⁵⁰

It has been pointed out that both versions of Pathfinder perform "well" when the domain expert himself enters observations.⁴⁹ This claim is presumably

based on the diagnosis of record as an appropriate "gold standard." Good performance in this situation is easily explained. Only a conditional probability of zero (or very close to zero) can eliminate competing diagnoses in a Bayesian model and drive the system to arrive at one "correct" diagnosis. The domain expert knows where the zeros are in the knowledge base since he determined the conditional probabilities. The expert diagnoses many lymph node sections quickly, using pattern recognition. When the expert is asked to identify the features that appear in such biopsy specimens, he is likely to report only features that confirm the already arrived-at diagnosis.⁴⁹ Thus, when entering findings into the Pathfinder/Intellipath system, the domain expert simply enters the features that best define a typical case, along with the pertinent negatives that the program needs to eliminate competing diagnoses.

One cannot criticize that approach since this is precisely how surgical pathologists generate typical microscopic descriptions. Busy pathologists who are technology oriented have a series of "boiler plate" descriptions already on a word processor. Microscopic descriptions (by the same individual) of different cases with the same final diagnosis end up being similar, if not identical, from case to case. This is because the description is that of a typical case, emphasizing selected features that define the entity. In our experience, based on a review of surgical pathology reports from tens of thousands of cases, microscopic descriptions may completely ignore features of a case that are not characteristic of the disease entity, no matter how prominent those features may be.

Conclusions

In our critique of Bayesian systems for diagnostic pathology, we have concentrated on a system for lymph node pathology, one of our areas of expertise. Lymph node biopsy interpretation is not an ill-structured task. The proper methods for processing and evaluating lymph node specimens are well known.^{51,52} In our experience, both in the United States and in Europe, the correct methods are not often followed, however. Lymph nodes need special processing, including unfixed tissue for immunologic studies, special methods of fixation, and technically optimal sections. Because of a lack of education and/or poor motivation, lymph nodes are often handled like other specimens in a busy surgical pathology laboratory, making diagnostic evaluation much more difficult. Careful attention to proper technique and the established methods of reasoning based on patterns will,

in our opinion, improve diagnostic accuracy and lead to better KBS.

In this article, we have asked many questions about the appropriateness of probabilistic systems for the problem-solving task of diagnosis. We certainly do not expect everyone to share our ideas or to agree that the evidence that we have cited is conclusive. Our goal is to ask penetrating questions that will stimulate further discussion about effective knowledge representation.

Our own view is that medical KBS development should be knowledge-oriented rather than technique-oriented. As Feigenbaum has pointed out, "the first principle of knowledge engineering is that the problem solving power exhibited by an intelligent agent's performance is primarily the consequence of its knowledge base."⁵³ Therefore, we believe that concentrating on the extraction and encoding of knowledge, rather than on the search for a particular type of inference mechanism to handle uncertainty, will result in more effective KBS.

We also view diagnostic KBS as important teaching tools. For this reason, we strive to develop diagnostic systems that use the same method of knowledge representation as an expert. In this way, any user who studies how the KBS arrive at interpretations learns a logical diagnostic approach to the domain. To enhance the educational value of our hematology workstation, we provide both an electronic textbook with an atlas of digitized photomicrographs,⁵⁴ and a series of case studies.⁵⁵

The Bayesian model and normative expert systems using that model are not usually designed with the goal of capturing knowledge in the form that domain experts teach and use in daily practice. It is, of course, perfectly justifiable to build an expert system that does not use the same knowledge representation as an expert, or that does not use knowledge in the same way the expert does, if such a system actually performs well.

Based on empirical results, Bayesian diagnostic systems in hematopathology have an accuracy rate of between 60% and 70% when the diagnosis of record is used as the "gold standard." Although some of the Bayesian system's misdiagnoses are relatively minor (such as a misclassification of one type of low-grade non-Hodgkin's lymphoma for another low-grade lymphoma) and, therefore, may not result in a change of therapy for the patient, we are simply not convinced that an overall diagnostic accuracy in the 60–70% range qualifies as performance that is at least as good as that of an expert. Furthermore, in our experience, the Bayesian diagnostic systems in he-

matopathology are capable of making major errors, such as diagnosing a malignant disease as a benign disease, with 100% certainty.

Many models of medical practice and diagnostic reasoning have been proposed. Some physicians simply do not believe in the distinctions between problem solving and decision making that have been proposed in the literature. In clinical medicine, problem solving and decision making are often viewed as being intertwined, as Elstein suggested.⁵ Since we are not clinicians, we make no attempt to argue that our view is appropriate for all of clinical medicine. We do know, however, that hematopathologists who face the problem of distinguishing malignant but potentially curable diseases of the hematopoietic system from benign "look-alikes" on a day-to-day basis do not view their role as that of a lottery player.

The methods that we are advocating for the creation of diagnostic KBS require that criteria be established for the domain. Some of the problems that can be encountered are related to the way that scientific evidence is gathered and used in medicine. A study of the agreement rate for differential diagnostic criteria between well-known experts in ovarian pathology demonstrated just how difficult it can be to acquire diagnostic criteria if the knowledge engineer is not experienced in the field.⁵⁶ In that study, eight American and European experts in ovarian pathology were given the name of an ovarian disease and asked to list all of the diagnoses that they considered to be morphologically similar to the given disease. There was very little consensus among the experts. The study concluded that once a pathologist leaves a training program, he or she develops his or her own diagnostic criteria through experience, and that the limited consensus in differential diagnosis can be traced back to a limited consensus in diagnostic criteria. The authors went on to say that a formal and explicit representation of knowledge, including pattern recognition techniques, can stimulate refinement of diagnostic criteria, and that these essential steps for improving diagnostic performance are applicable to other domains in medicine, as well.⁵⁶

Admittedly, it is difficult to find knowledge engineers who have sufficient experience in a particular specialty to make the domain expert concentrate on those particular features that define a pattern or disease, rather than on all of the possible features that can be associated with a given disease. From our point of view, however, choosing an easy to implement but conceptually "incorrect" model for KBS design will not lead to effective, usable systems in those situations where accurate diagnosis is paramount.

References ■

1. Lipkin M, Engle RL Jr, Davis BJ, et al. Digital computer as aid to differential diagnosis: use in hematologic diseases. *Arch Intern Med.* 1961;108:56-72.
2. Coiera E. Question the assumptions. In: Barahona P, Christensen JP, eds. *Knowledge and Decisions in Health Telematics.* Amsterdam: IOS Press, 1994:61-6.
3. van der Lei J. Computer-based decision support: the unfulfilled promise. In: Barahona P, Christensen JP, eds. *Knowledge and Decisions in Health Telematics.* Amsterdam: IOS Press, 1994:67-72.
4. Cooper G. Probabilistic and decision-theoretic systems in medicine. *Artif Intell Med.* 1993;5:289-92.
5. Elstein AS. Paradigms for research on clinical reasoning: a researcher's commentary. *Teach Learn Med.* 1992;4:147-9.
6. Sultan C, Imbert M, Prioret G. Decision-making system (DMS) applied to hematology: diagnosis of 180 cases of anemia secondary to a variety of hematologic disorders. *Hematol Pathol.* 1988;2:221-8.
7. Heckerman DE, Horvitz EJ, Nathwani BN. Toward normative expert systems: Part I. The Pathfinder project. *Methods Inf Med.* 1992;31:90-105.
8. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. *Science.* 1959;130:9-21.
9. Deber RB, Baumann AO. Clinical reasoning in medicine and nursing: decision making versus problem solving. *Teach Learn Med.* 1992;4:140-6.
10. Gage JS. Notes from Stony Brook. *MD Comput.* 1993;10:370-2.
11. Diamond LW, Nguyen DT. Interpretative reporting. *Lab Med.* 1993;24:530.
12. Szolovits P, Pauker SG. Categorical and probabilistic reasoning in medical diagnosis. *Artif Intell.* 1978;11:115-44.
13. Bartels PH. The diagnostic pattern in histopathology. *Am J Clin Pathol.* 1989;91(Suppl 1):S7-S13.
14. Nguyen DT, Diamond LW, Prioret G, Sultan C. Expert system design in hematology diagnosis. *Methods Inf Med.* 1992;31:82-9.
15. Clancey WJ. Heuristic classification. *Artif Intell.* 1985;27:289-350.
16. Nguyen DT, Diamond LW, Prioret G, Sultan C. A decision support system for diagnostic consultation in laboratory hematology. In: Lun KC, Degoulet P, Piemme TE, Rienhoff O, eds. *MEDINFO 92.* Amsterdam: North Holland, 1992:591-5.
17. Diamond LW, Nguyen DT, Sheridan BL, Strul M, Bailey K, Bak A. An automated hematology laboratory with computer-controlled robotics. In: Greens R, et al. *MEDINFO 95* (in press).
18. Nguyen DT, Diamond LW. The concept of a computer workstation for diagnostic haematology. *Aust J Med Sci.* 1993;14:71-5.
19. Diamond LW, Nguyen DT, Jouault H, Imbert M. Evaluation of a knowledge-based system for interpreting flow cytometric immunophenotyping data. In: Reichert A, Sadan BA, Bengtsson S, Bryant J, Piccolo U, eds. *Proceedings of MIE 93.* London: Freund, 1993:124-8.
20. Diamond LW, Nguyen DT, Andreeff M, Maiese RL, Braylan RC. A knowledge-based system for the interpretation of flow cytometry data in leukemias and lymphomas. *Cytometry.* 1994;17:266-73.
21. Nguyen DT, Cherubino P, Tamino PB, Diamond LW. Computer-assisted bone marrow interpretation: a pattern approach. In: Reichert A, Sadan BA, Bengtsson S, Bryant J, Piccolo U, eds. *Proceedings of MIE 93.* London: Freund, 1993:119-23.
22. Nguyen DT, Diamond LW, Cherubino P, Koala WBJ, Imbert M, Andreeff M. A diagnostic workstation for neoplastic bone marrow diseases: evaluation on 526 cases. In: Greens R, et al. *MEDINFO 95* (in press).
23. Nguyen DT, Mishka VG, Moskowitz FB, Diamond LW. A multiparameter decision support system for lymph node pathology. In: Barahona P, Veloso M, Bryant J, eds. *Proceedings of MIE 94.* Lisbon: 1994:162-7.
24. Diamond LW, Nguyen DT. Communication between expert systems in haematology. In: Richards B, ed. *Healthcare Computing 1993: Current Perspectives in Healthcare Computing.* Weybridge, UK: BJHC Books, 1993:111-9.
25. Diamond LW, Mishka VG, Seal AH, Nguyen DT. A clinical database as a component of a diagnostic hematology workstation. In: *Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care.* J Am Med Informatics Assoc Symposium Suppl. 1994;1,suppl:298-302.
26. Feinstein AR. An analysis of diagnostic reasoning: I. The domains and disorders of clinical microbiology. *Yale J Biol Med.* 1973;46:212-32.
27. Feinstein AR. An analysis of diagnostic reasoning II. The strategy of intermediate decisions. *Yale J Biol Med.* 1973;46:264-83.
28. Feinstein AR. An analysis of diagnostic reasoning III. The construction of clinical algorithms. *Yale J Biol Med.* 1974;47:5-32.
29. Warner HR, Haug P, Bouhaddou O, et al. ILIAD as an expert consultant to teach differential diagnosis. In: *Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press, 1987:371-6.
30. de Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis in acute abdominal pain. *BMJ.* 1972;2:9-13.
31. Leaper DJ, Horrocks JC, Staniland JR, de Dombal FT. Computer-assisted diagnosis of abdominal pain using "estimates" provided by clinicians. *BMJ.* 1972;4:350-4.
32. Patil RS, Szolovits P, Schwartz WB. Modeling knowledge of the patient in acid-base and electrolyte disorders. In: Szolovits P, ed. *Artificial Intelligence in Medicine.* Boulder, CO: Westview Press, 1982:191-226.
33. Pauker SG, Kassirer JP. Decision analysis. *N Engl J Med.* 1987;316:150-7.
34. Shortliffe EH, Buchanan BG. A model of inexact reasoning in medicine. *Math Biosci.* 1975;23:351-79.
35. Gordon J, Shortliffe EH. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artif Intell.* 1985;26:323-57.
36. Zadeh L. The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems.* 1983;11:199-227.
37. Kingsland LC III, Sharp GC, Kay DR, Weiss SM, Roeseler GC, Lindberg DAB. An expert consultant system in rheumatology: AI/RHEUM. In: *Proceedings of the Sixth Annual Symposium on Computer Applications in Medical Care.* 1982:748-52.
38. Lipkin M, Hardy JD. Mechanical correlation of data in differential diagnosis of hematological diseases. *JAMA.* 1958;166:113-23.
39. Alvey PL, Myers CD, Greaves MF. High performance for expert systems: I. Escaping from the demonstrator class. *Med Inf (Lond).* 1987;12:85-95.
40. Alvey PL, Preston NJ, Greaves MF. High performance for expert systems: II. A system for leukaemia diagnosis. *Med Inf (Lond).* 1987;12:97-114.
41. Diamond LW, Mishka VG, Seal AH, Nguyen DT. Multiparameter interpretive reporting in diagnostic laboratory hematology. *Int J Biomed Comput.* 1994;37:211-24.
42. Imbert M, Prioret G, Dadi W, Sultan C. An expert system

- applied to the diagnosis of anemia with special reference to myelodysplastic syndromes. *Blood Cells*. 1989;15:563-71.
43. Heckerman D, Horvitz EJ, Nathwani BN. Toward Normative Expert Systems: The Pathfinder Project. Knowledge Systems Laboratory Report KSL-90-08. Stanford, CA: Stanford University, 1990.
 44. Nathwani BN. Diagnostic significance of morphologic patterns in lymph node proliferations. In: Knowles DM, ed. *Neoplastic Hematopathology*. Baltimore: Williams & Wilkins, 1992:407-25.
 45. Nathwani BN, Burke JS, Winberg CD. Architectural features of normal, neoplastic, and non-neoplastic lymph nodes: a practical diagnostic approach. In: Murphy GF, Martin MC, eds. *Lymphoproliferative Disorders of the Skin*. Boston: Butterworth, 1986:73-119.
 46. Nathwani BN, Brynes RK, Lincoln T, Hansmann ML. Classifications of Non-Hodgkin's lymphomas. In: Knowles DM, ed. *Neoplastic Hematopathology*. Baltimore: Williams & Wilkins, 1992:555-601.
 47. Nathwani BN, Heckerman DE, Horvitz EK, Lincoln TL. Integrated expert systems and videodisc in surgical pathology: an overview. *Hum Pathol*. 1990;21:11-27.
 48. Rolston DW. *Principles of Artificial Intelligence and Expert Systems Development*. New York: McGraw-Hill, 1988.
 49. Heckerman DE, Nathwani BN. An evaluation of the diagnostic accuracy of Pathfinder. *Comput Biomed Res*. 1992;25:56-74.
 50. Li Y-C, Haug PJ. Evaluating the quality of a probabilistic diagnostic system using difference inference strategies. In: *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1993:471-7.
 51. Knowles DM. Organization and operation of a hematopathology laboratory. In: Knowles DM, ed. *Neoplastic Hematopathology*. Baltimore: Williams & Wilkins, 1992:323-66.
 52. Banks PM. Technical factors in the preparation and evaluation of lymph node biopsies. In: Knowles DM, ed. *Neoplastic Hematopathology*. Baltimore: Williams & Wilkins, 1992:367-84.
 53. Luger GF, Stubblefield WA. *Artificial Intelligence Structures and Strategies for Complex Problem Solving*, 2nd ed. Redwood City, CA: Benjamin/Cummings, 1993:308.
 54. Diamond LW, Nguyen DT. Linking an electronic textbook to expert systems in haematology. In: Richards B, ed. *Healthcare Computing 1994: Current Perspectives in Healthcare Computing*. Weybridge, UK: BJHC Books, 1994:269-79.
 55. Diamond LW, Tamino PB, Seal AH, Nguyen DT. Multiparameter case studies using knowledge-based systems in hematology. In: Barahona P, Veloso M, Bryant J, eds. *Proceedings of MIE 94*. Lisbon: 1994:168-73.
 56. van Ginneken, van der Lei J. Understanding differential diagnostic disagreement in pathology. In: *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1991:99-103.