A.M. van Ginneken

Institute of Medical Informatics,
Erasmus University Rotterdam,
The Netherlands

Commentary

Diagnostic Support: Towards the Intelligent Integrated Reference Source

Reflections on R.A. Miller et al.'s paper:
INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine.

Introduction

Since the 1970s, diagnostic expert systems have been an important area of research in medical informatics [1]. Researchers had high expectations of artificial intelligence as an aid in medical diagnosis. They aimed at making expertise available to a much larger medical community than personal consultation of experts would allow. Incentives were the increasing difficulty of keeping up-to-date with state-of-the-art medical knowledge, the ability of computers to rapidly process a large number of possibilities, and the absence of bias due to human factors. Examples of human factors are: previous training and experience, recently encountered diseases, fatigue, and time pressure.

Some diagnostic expert systems perform comparably to experts. *Internist-I and,* its successor QMR (Quick Medical Reference), DxPLAIN, ILIAD, and MEDITEL have become widely known for their scope and quality [2]. Several systems have become a commercially available. Yet, no large-scale diagnostic expert system has found widespread use in the field of internal medicine and many other applications never progressed beyond an experimental stage.

This paper summarizes why diagnostic expert systems are not yet widely used and discusses shifts in both the type and methodology of diagnostic decision support.

Early Experiences with *Internist-I*

The evaluation of *Internist-I* [3] produced a number of valuable insights into the strengths and weaknesses of the application. As opposed to MYCIN [4], of which the rules combine declarative and procedural knowledge, *Internist-I* had the advantage that its knowledge base contained purely declarative knowledge in the form of intuitively appealing disease profiles. Acquisition of purely declarative knowledge is more transparent and easier to manage when the system scales up. As a result, maintenance of the knowledge base (KB) was more straightforward and could be done by medical experts with little need of a knowledge engineer. The numeric attributes of each manifestation in a disease profile were innovative and intuitive, and circumvented the problems of probabilistic approaches with respect to normalization, a posteriori probabilities, and interdependencies. Despite these strengths, the limitations of *Internist-I* mainly involved its scope, explanation capabilities, and lack of integration in clinical practice. The limitations that inspired and influenced subsequent research efforts will be discussed.

Scope and Transparency

A well-known problem of diagnostic expert systems is their limited scope. Because knowledge acquisition and maintenance is an ongoing labor-intensive process, most systems are restricted to small domains. To cover all of internal medicine is a laudable goal. Although *Internist-I* had a large KB, the developers were aware that it did not cover all diseases relevant to the domain of internal medicine. To give it a "fair" test, the evaluation of the system was limited to diseases that were covered by its KB [2]. Hence, the problem remains to decide when a system is ready for use in the non-restricted world of reality. More important than being "fair" for the system, a test must also be "fair" for the user. Part of the problem is that the system does not know what it does not know: it cannot warn the user that it did not consider disease A, because disease A is not covered by its KB. The other, even more important, part of the problem is that the boundaries of the system's contents and capabilities are rarely evident for the user. For example, if a differential diagnosis does not include disease A it is not immediately evident whether the system rejected disease A or was not even capable to consider it.

In *Internist-I*, domain coverage is explicitly represented in its disease profiles. Yet, a paradox remains: if the user can reliably determine that the KB covers his problem area, he does not need the system to form his problem area. And if the user cannot determine if the KB covers his problem area, how well can the system help him?

In this respect it is interesting to note that books are well accepted as reference sources, despite their limited scope. No one will say that it is dangerous to use a book, because the book does not know what it does not cover. If the reader finds in a book what he looks for, it belongs to the scope of the book, otherwise it doesnot. In analogy, the perceived reliability of an expert system strongly depends on the transparency of that system.

Integration

Like most other diagnostic expert systems, *Internist-I* was a stand-alone system. The interactive sessions with the system required re-entry of data and were time consuming. This largely precluded the usability of *Internist-I* during routine patient care. Integration of diagnostic expert systems in patient care means that such systems use the information routinely collected during the health-care process. It is obvious that patient data have to be electronically available in an interpretable format to make such integration possible. Support of structured data entry (SDE) by clinicians is an important area of current research [5-7]. But even if the record provides sufficient structure for integration with a diagnostic support system, several problems remain. Semantic mapping between the vocabulary of the CPR and that of the expert system may be difficult because of partial matches. Furthermore, interaction with the diagnostic support system will be required to refine the initial differential diagnosis. This interaction may elicit additional data from the clinician, but these data cannot automatically be recorded in the record: sometimes data are entered to direct the reasoning process without reflecting actual patient findings. Consequently, data generated during interaction with a diagnostic support system must be verified by the user before they can be added to the CPR [8]. Technical and semantic challenges are not the only factors that inhibit integration. In contrast with primary care in the UK and the Netherlands [9], specialized care is still predominantly paper based.

Another drawback of a stand-alone system is that its use depends entirely on the initiative of the user. If users feel comfortable about a diagnosis they will not consult reference knowledge, but such a feeling does not guarantee a correct diagnosis.

Explanation

Reference knowledge as laid down in books is diagnosis-oriented, whereas medical diagnosis is findings-oriented. Hence, the organization of knowledge in books is not optimal for diagnostic problem solving. Findings-oriented organization of knowledge on paper would result in unacceptable redundancy. Furthermore, paper is passive and cannot actively help to direct the reader to the answer. Computers, on the other hand, can organize a set of data items in many different ways and apply reasoning algorithms to its contents.

Physicians, however, do not seek a "black box" that produces a diagnosis without further explanation [10,11]. Such a black box would unjustly discard the expertise of the clinician and may be experienced as de-skilling [12]. At the time of the evaluation of *Internist-I*, the authors were aware of its limited explanation capabilities. The system was unable to distinguish causal from predisposing factors, and findings explained by a diagnosis were no longer used to evoke new disease hypotheses. Furthermore, interdependent manifestations disproportionally favored the most common disease that explained them.

Hence, the developers of *Internist-I* were very aware that clinicians want a consultant to whom they can ask directed questions. The answers will supplement their own knowledge and need to solve the missing links in their line of reasoning.

Quick Medical Reference (QMR)

QMR is the successor of *Internist-I* and it has eliminated many of *Internist-I*'s shortcomings [13]; especially its explanation capabilities have been greatly improved. Examples are the explicit representation of causal and predisposing factors, improved handling of interdependencies, and the option for the user to "keep" already explained findings for the creation of new disease hypotheses. In addition, QMR supports a wide variety of questions to aid clinicians in solving problems. Among others, the physician can ask for a finding "work-up", ask for questions to rule-in or rule-out a diagnosis, compare diagnoses, and critique a user-specified disease hypothesis in the context of the current case. Apart from these powerful diagnostic aids, the user can easily view the KB contents: browse the disease and finding indexes, and view individual disease profiles.

With its extended KB, QMR has become a powerful diagnostic aid. It is not intended to replace the physician, but to serve as an intelligent reference source. At this moment, lack of integration in routine health care is still a major impediment to the use of diagnostic expert systems.

Evolution of Decision-Support Systems

Experiences with early expert systems have initiated a variety of new developments and research efforts. Early developments were for a large part driven by what information technology *could* make possible; emphasis was on the potential and validity of a methodology. Later developments were also driven by what is *feasible*. In the evolution of decision-support systems three main areas deserve attention: a shift in the focus of decision support, exploration of new methods for inferencing, and changes in maintenance policy.

Shift in Focus

The initial euphoria about diagnostic expert systems dwindled when information technology failed to become widely used for the collection of patient data. Medical diagnosis is a complex process, often requiring a large data set of unpredictable scope. Lack of structured and standardized data in electronic format precluded integration of diagnostic expert systems in routine health care. Consequently, the focus of medical decision support shifted to what the infrastructure would allow. The first data that became available in a structured format included demographics, vital signs, and laboratory results. These data were used by alerting and reminder systems to draw clinicians' attention to undesirable or even dangerous findings [14,15]. Ongoing developments in medical record applications included drug prescriptions and, especially in primary care, problem-oriented recording and SOAP codes in the clinical narrative [16]. Coding in ICD, ICPC, or SNOMED further improved the interpretability of patient data [17]. Gradually, the contents of patient records permitted automated generation of useful remarks to assist physicians with specific well-defined tasks, such as protocol-based care and adherence to practice guidelines. The so-called "critiquing" systems became an important area of development in medical decision support [18-20].

The data needed for diagnostic support are partly hidden in free-text encounter notes, radiology, pathology, and other narrative reports. Automated interpretation of such texts still does not compete with the clinician. Developments in the acquisition of structured data and increased efforts for standardization of patient records constitute important steps towards the feasibility of integrated diagnostic support.

Shift in Methodology

Researchers in the field of diagnostic decision support have always struggled with the quantification and formalization of the complex diagnostic process. Quantitative approaches faced the problem of lacking probabilistic data: a priori probabilities were incomplete and not normalized, a posteriori probabilities were largely absent, and interdependencies difficult to quantify. Bayesian applications often used simplified models, ignoring normalization and interdependencies. Experiments were designed to adjust probabilities in an attempt to achieve expert performance. It is noteworthy, however, that ''simple'' Bayes with the assumption of independence tends to perform better than ''proper'' Bayes, without that assumption, if the required parameters cannot be reliably derived from extensive patient data sources [21].

Other systems based their inference engines on models reasoning with hypothesis sets, such as fuzzy logic and Dempster-Shafer [22]. MYCIN and *Internist-I* used heuristic approaches: MYCIN used Certainty Factors to express belief in rules and *Internist-I* and the early QMR used intuitive expressions for sensitivity and specificity in the form of ''frequency'' and "evoking strength", respectively.

The qualitative aspects of the KBs were necessary to provide adequate explanation facilities and insight in the contents of the knowledge base. Maintenance of the KB required extensive literature study and consultation of experts. If the collection of qualitative knowledge was a big task, so much more was the acquisition and maintenance of the quantitative parameters.

In an era of labor-intensive knowledge engineering, the burden of KB maintenance sparked a marked interest in neural-net technology [23-25]. The learning capabilities of neural-net applications led to high expectations with respect to the effort needed to create and maintain systems with expert performance. Several systems did indeed perform comparably to experts, but this success was not observed in domains where explanation capabilities were important. Neural-net technology applied to medical diagnosis produced unsatisfactory "black box"' output [26].

Quantitative developments continued. Bayesian belief networks allow for the expression of interdependencies, and the partitioning of the network in local belief networks permits an intuitive acquisition of the required quantitative parameters [27,28]. QMR-DT was a successor of QMR, based on a probabilistic model [29]. The qualitative content of the KB, in combination with the far more transparent quantitative parameters, caused increasing interest in Bayesian belief networks.

Quantitative aspects of diagnostic reasoning, such as the ranking of disease hypotheses, are not equally useful for each user. Given a set of findings, a general practitioner will act strongly on disease prevalence, whereas a histopathologist seeks the best morphologic match. From primary towards tertiary care, or from generalist to expert, diagnostic reasoning changes from prevalence-driven to pattern-driven.

Shift in Maintenance

Neural nets provide no solution in areas where the user requires insight into the reasoning process. Hence, acquisition and maintenance of KBs for medical decision support remain an arduous undertaking. Research groups rarely have the resources to undertake this task beyond the demands of an experimental setting. An average disease profile in QMR is often based on more than 70 publications [30]. Established professional organizations are most suited to take on the effort and responsibility to provide a KB with sufficient continuity and scope for use in clinical practice. Primary care organizations have established task forces to create practice guidelines. Similar activities in specialized health care produce dedicated treatment protocols. The combined effort of system developers and professional organizations offers potential for integrated and reliable decision support. It is, however, important that partners take responsibility for their own share in the expertise. In other words, responsibilities for software and content are preferably separated.

The Future of Diagnostic Decision Support.

Past experience has taught several important lessons. First, optimal benefit of decision support requires integration with routine care. Second, the scope and reasoning process of decision-support systems has to be transparent. And third, maintenance of the KBs for decision-support systems has to be taken up by professional organizations.

What does this mean for diagnostic decision support? It is obvious that integration of such complex functionality requires a well-structured CPR with a widely accepted vocabulary. But given such a CPR, medical diagnosis is so complex that an expert system with the ideal scope and coverage is an unrealistic aim. The future of diagnostic expert systems lies in two types of functionality, which can well be combined. One is the "watchdog", which will only alert clinicians when their diagnostic hypotheses lack a probable diagnosis or are in conflict with certain observations. The philosophy of the watchdog is that it supplements the clinician's own thinking: each remark is a bonus. Potential pitfalls, however, are disregard of the comments due to poor specificity, or too much reliance on the watchdog. The other type of functionality is the "intelligent" reference source, which can dynamically generate views tailored to specific clinical questions. In the future, CPR applications will inevitably become more rule than exception in patient care. They will pave the way for the descendents of *Internist-I*.

References

1. Miller RA. Medical diagnostic decision support systems – past, present, and future: a threaded bibliography and brief commentary. J Am Med Inform Assoc 1994;1:8-27.
2. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. N Engl J Med 1994;330:1792-6.
3. Miller RA, Pople HE, Myers JD. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 1982;307:468-76.
4. Shortliffe EH. *Computer-based Medical Consultations: MYCIN.* New York: Elsevier/North Holland, 1976.
5. Trace D, Naeymi-Rad F, Haines D, et al. Intelligent Medical Record--entry (IMR-E). J Med Syst 1993;17:139-51.
6. Howkins TJ, Kay S, Rector AL, et al. An overview of the PEN & PAD project. In: Rienhoff O, Lindberg DAB, eds. *Lecture Notes in Medical Informatics. No. 40.* Berlin: Springer Verlag, 1990:73-78
7. 7.Moorman PW, Van Ginneken AM, Van der Lei J, Van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Meth Inform Med 1994;33:454-63.
8. Van Ginneken AM, Liem EB, Moorman PW. Integrating QMR with a computer-based patient record. In: Safran C. *Proceedings of the 17th SCAMC.* New York: McGraw-Hill Inc, 1993:98-102.
9. Van der Lei J, Duisterhout JS, Westerhof HP, et al. The introduction of computer-based patient records in The Netherlands. Ann Intern Med 1993;119:1036-41.
10. Miller RA, Masarie FE. The demise of the 'Greek oracle' model for medical diagnostic systems. Meth Inform Med 1990;29:1-2.
11. De Dombal F.T. Editorial Commentary. Computer-aided diagnosis and medical decision support are not synonymous. Meth Inform Med 1995;34:369-70.
12. Heathfield H, Kirkham N. A cooperative approach to decision support in the differential diagnosis of breast disease. Med Inf 1992;17:21-33.
13. Parker RC, Miller RA. Creation of realistic appearing simulated patient cases using the INTERNIST-I/QMR knowledge base and interrelationship properties of manifestations. Meth Inform Med 1989;28:346-51.
14. McDonald CJ, Hui SL, Smith DM, et al. Reminders to physicians from an introspective computer medical record. Ann Intern Med 1984;100:138.
15. Kuperman GJ, Gardner RM, Pryor TA. *HELP: A dynamic hospital information system.* New York: Springer Verlag, 1991.
16. Van Ginneken AM. The structure of data in medical records. Review paper. In: Van Bemmel JH, McCray AT. *Yearbook of Medical Informatics*, Stuttgart: Schattauer Verlag, 1995:61-70.
17. Cimino JJ. Review Paper: Coding Systems in Health Care. Meth Inform Med 1996;35:273-84.
18. Van der Lei J, Musen MA, Van der Does E, et al. Comparison of computer-aided and human review of general practitioners' management of hypertension. Lancet 1991;338:1505-8.
19. Miller PL, Frawley SJ. Trade-offs in producing patient-specific recommendations from a computer-based clinical guideline: a case-study. J Am Med Inform Assoc 1995;2:238-42.
20. Grimshaw J, Freemantle N, Wallace S, et al. Developing and implementing clinical practice guidelines. Qual Health Care 1995;4:55-64.
21. Gammerman A, Thatcher AR. Bayesian diagnostic probabilities without assuming independence of symptoms. Meth Inform Med 1991;30:15-22.

22. Van Ginneken AM, Smeulders AWM. An analysis of five strategies and their suitability in pathology. Anal Quant Cytol Histol 1991;13:93-109.
23. Hertz J, Krogh A, Palmer RG. Introduction to the theory of neural computation. Redwood City CA: Addison-Wesley Publishing Co, 1991.
24. Penny W, Frost D. Neural networks in clinical medicine. Med Decis Making. Review. 1996;16:386-98.
25. Forsstrom JJ, Dalton KJ. Artificial neural networks for decision support in clinical medicine. Review. Ann Med 1995;27: 509-17.
26. Hart A, Wyatt J. Evaluating black boxes as medical decision aids: Issues arrising from a study of neural networks. Med Inform 1990;15:229-36.
27. Herskovits EH, Copper GF. Algorithms for Bayesian belief-network precomputation. Meth Inform Med 1991;30:81-9.
28. Heckerman DE, Nathwani BN. Toward normative expert systems: Part II. Probability-based representations for efficient knowledge acquisition and inference. Meth Inform Med 1992;31:106-16.
29. Shwe MA, Middleton B, Heckerman DE, et al. Probabilistic diagnosis using a reformulation of the INTERNIST-I/QMR knowledge base I. The probabilistic model and inference algorithms. Meth Inform Med 1991;30:241-55.
30. Giuse NB, Giuse DA, Miller RA, et al. Evaluating consensus among physicians in medical knowledge base construction. Meth Inform Med 1993;32:137-45.

Address of the author:
Astrid M. van Ginneken, MD, Ph.D.,
Institute of Medical Informatics,
Erasmus University Rotterdam,
PO box 1738,
3000 DR Rotterdam,
The Netherlands.
E-mail: vanginneken@mi.fgg.eur.nl