M. Fieschi

Service de l'Information Médicale,
Hôpital de la Timone A,
Marseille, France

Synopsis

Decision-Support Systems

This section of the Yearbook includes five articles that represent the scope and range of developments of *Decision-Support Systems* in the field of medical informatics. Because medical practice implies medical decision making, health professionals have tried, from the very beginning of the computer age, to develop advisory systems that would assist in the diagnostic/therapeutic process. The methods of knowledge representation, reasoning, management of uncertainty, and the strategies of using the knowledge bases have stimulated scientific reflection (and also sometimes controversies) from the beginning. The number of publications on such developments is impressive but their quality does not always correspond with potential "usability" in medical practice. Several barriers limit their effective implementation. Excellent decision-making capabilities do not guarantee system utility and successful implementation. For example, as everyone knows, it is not realistic to view diagnosis as separated from the process of deciding what additional information would be most useful and what is the best option for therapy management. The designers of an advisory system should take into account this "usability context".

Nevertheless, today we witness the emergence of commercial versions of computer decision-support programs (e.g., QMR [1], DXplain [2], Illiad [3]) but they are still used mainly in educational settings. Though the systems and methodologies developed cover a broad range of problems and medical applications, few such systems are in routine clinical use. Many reasons have been given to explain this frustrating situation. Two of the selected papers [4-5] deal with some of these reasons. They present, from different points of view, a reflection on thirty years of research in the field, analyze the current state of clinical decision-support systems, and explain their limitations in producing tools for improving health care. In the recent past, the central question was: How to achieve the functionality of decision-support tools? Now, the question has shifted to: What functionalities should such systems have? These articles propose research perspectives and coherent approaches to the development of such tools. In my opinion, it is important that researchers in the domain consider the arguments of these excellent analyses that restate the medical decision-support approach, in its context and complexity. A third article [6] presenting an evaluation of explanations of probabilistic inference, has also a rationale linked to the acceptability of decision-support systems for clinical use. The last two papers present applications of neural networks. These techniques offer several advantages and attractions. The most important are: the suitability for machine learning, their powerful model of knowledge representation, the inherent parallelism of the algorithms and their robust behavior with noisy data. Many researchers in the decision-support field are enthusiastic about these techniques as they have now been extensively applied over five years in medicine as well as in other fields. The two applications presented here were developed in radiology [7-8].

The first paper is by Shortliffe: *The adolescence of AI in medicine: Will the field come of age in the '90s?* [4]. It presents a self-examination of the Artificial Intelligence in Medicine (AIM) community, a discussion and personal views that concern "soul-searching issues" for AI in medicine. The author analyses the following paradoxical situation: on the one hand the significant contribution of medical AI researchers to the AI field and computer science is established. On the other hand, the medical community still resists the adoption and use of expert-systems techniques. According to Shortliffe (and I fully agree with his views), the problem has little to do with AI and the quality of AIM research. Resistance to system use arose despite the quality of the AI methods developed and applied in the field. Since AIM is one of the core areas of *medical informatics* - itself an interdisciplinary field - the context in which AIM concepts will ultimately be delivered, i.e., as useful tools integrated in medical practice, is essential. It helps to understand the complexity of the problems and to define the range of issues that AIM researchers need to consider. Because we must design and build information architectures that support, advance and promote better and more efficient health care, the AIM researchers have to enlarge their expertise beyond medical AI to other critical topics. For example, classical statistics, human-computer interaction, databases, information retrieval, information systems, standards for communications and knowledge exchange condition the effective delivery of decision-support systems in integrated professional environments. This is one of the main recommendations of this excellent and thoughtful paper.

Heathfield and Wyatt published *Philosophies for the design and development of clinical decision-support systems* [5]. This paper is, in a way, complementary to the first one. It is a lucid discussion which highlights that a philosophy of "how" to build clinical decision-support systems is central to the successful development of such systems and to the discipline of medical informatics. The authors analyze the current state of clinical decision-support systems, the motivations of developers and the perceptions of potential end-users. They underline some factors that may impede the successful development of such systems. Among many factors that can negatively affect the expansion of a development philosophy, the paper addresses four particularly significant factors :

- A preoccupation with computer artifacts (for example, the choice of the software tool is very often the first choice of the project) dictates the approach of the decision problem.
- The failure to adopt appropriate models for analyzing a problem causes, like computer artifacts, "tunnel vision" and impedes the research of more appropriate solutions.
- A lack of clear language for communicating the development philosophy and facilitating better communication and understanding of the problems between the clinician and the designer.
- A disregard for organizational issues. The authors stress the importance of the stability of a project workforce.

The example of the development of the ACORN system, and the analysis of the difficulties encountered, illustrate their arguments very well. In the last section, the authors describe a development philosophy that embodies many of their beliefs concerning the development of clinical decision-support systems. They define a development philosophy as a shared perception on how a problem or a class of problems should "in principle", be solved. They propose to establish and clarify the need to use a coherent methodology for modelling the problem through rapid prototyping, and to use appropriate methods and tools. They strongly advocate the need for evaluation and for a professional approach to implementation, maintenance and support.

Suermondt and Cooper published *An evaluation of explanations of probalistic inference* [6]. The aim of this contribution is to explain the conclusions of a probabilistic decision-support system. As has been highlighted by several authors, one of the prominent requirements given by practitioners for the acceptability of such systems in routine practice is the ability of the system to explain its advice. The article presents the results of a study in which the effects of the resulting explanations on users in the domain of anesthesia has been tested using the INSITE system. They compare subjects who had access to a belief network with explanations of the inference results to control subjects using the system without explanations.

The result of this study indicate that the addition of explanations to the system improved subjects' diagnostic accuracy. More precisely, the subjects who received explanations avoided adding incorrect diagnoses to their differential diagnosis more often than did subjects who received no explanations. However, explanations did not substantially affect whether the users rated the computer advice as "helpful" versus "useless". The authors recognize that, due to pragmatic constraints on the study, they did not have an opportunity to fully determine the areas in which explanations of a belief-network advice can have clinical utility. The paper is interesting but the complex question of the added value and the necessity of explanations in a decision-support system in practice is still open. Among the questions that remain: Does the explanation allow the user to eliminate the incorrect advice? Does the impact of explanations on diagnostic accuracy depend on the clinical domain? In a real clinical environment, would clinicians want to use the explanation facility?

The two last papers of this section report on neural network applications. These techniques have recently gained much attention for diagnostic tasks in medical imaging. The articles present applications in this field.

The first paper is by Tourassi and Floyd: *Artificial neural network for single photon emission computer tomography. A study of cold lesion detection and localization* [7]. The diagnostic task of the study includes uncertainty about the number of lesions present, their locations, sizes and the image noise level. The neural network, developed for lesion detection task in SPECT images, had three fully interconnected layers (an input layer with 256 nodes, a hidden layer with eight nodes - the number was determined experimentally - and an output layer with a single "decision" node). The trained network was evaluated on two classes of new images: those containing lesion sizes and noise levels which were included in the training set and those containing noise levels and which reconstructed with filters not included in the training set. The diagnostic performance of the system was evaluated at two noise levels using the free-response operating characteristic (FROC) analysis. The results showed high sensitivity and a low false-positive rate per image for all test situations.

Wu, Giger et al. reported in a study entitled *Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer* [8], the potential utility of neural networks as a decision-making aid to radiologists in the analysis of mammographic data.

The three-layer network used was trained to distinguish between benign and malignant lesions with 43 image features extracted from mammograms by experienced radiologists. Performance of the neural networks in classifying lesions was evaluated with ROC analysis. The results presented indicate that the patterns learned, with a jack-knife method on subsets of 132 cases from a training database of 133 cases, served as a good basis for classifying the remaining case. The effects of observer variation in feature ratings and reduction of the number of input features have been investigated. The study compared the performance of junior and senior mammographers with that of the neural network. Both were evaluated in classifying lesions as benign or malignant, using the same clinical cases. The performance of the neural network was higher than the average performance of the attending radiologists and residents, and the system seems to be able to recommend an appropriate action (biopsy or follow-up).

A last comment refers to the methodological problems of evaluation. Many alternative methods have been proposed for verification, validation and user acceptance testing. Each of these methods has its advantages and disadvantages. Various practical considerations enter into decisions concerning which evaluation method should be employed. The latter papers on the applications in radiology illustrate the difficulty in designing a protocol evaluation following the development stages of the system, and in conducting unbiased assessment studies. The independence between developers of the system and evaluators is necessary, as well as a large enough unbiased sample of clinical cases and the clear definition of the Gold Standard. But the impact of the system on users and health care is, unfortunately, poorly or not at all addressed in most of the evaluation studies.

In conclusion, the papers presented reflect the problematic nature of decision-support systems today and make significant contributions to further developments in clinical support improvement.

References

1. Miller RA, McNeil MA, Challinor SM, Maserie FW, and Myers JD. The Internist/Quick Medical Reference project status report. West J Med 1986;145:816-22.
2. Barnett GO, Cimino JJ, Hupp JA, and Hoffer EP. DXplain: An evolving diagnostic decision support system. JAMA 1987;258:67-74.
3. Sorenson D, Cundick RM, Fan C, and Warner HR. Passing partial information among Bayesian and Boolean frames. In: *Proc. 13th Annual Symp on Computer Applications in Medical Care*. Washington DC, 1989:50-3.
4. Shortliffe EH. The adolescence of AI in medicine: Will the field come of age in the '90s? Artif Intell Med 1993;5:93-106
5. Heathfield HA, and Wyatt J. Philosophies for the design and development of clinical decision-support systems. Methods Inf Med 1993;32:1-8
6. Suermondt HJ, and Cooper GF. An evaluation of explanations of probabilistic inference. Comput Biomed Res 1993;26:242-54.
7. Tourassi GD, and Floyd CE. Artificial neural networks for single photon emission computed tomography. A study of cold lesion detection and localization. Invest Radiol 1993;28:671-7.
8. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, and Metz CE. Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. Radiology 1993;187:81-7.

Address of the author:
M. Fieschi,
Service de l'Information Médicale,
Hôpital de la Timone A,
Rue Saint Pierre,
13385 Marseille Cedex 5,
France