Viewpoint ■

The Practice of Informatics

Artificial Intelligence in Medicine: The Challenges Ahead

ENRICO W. COIERA, MB, PHD

Abstract The modern study of artificial intelligence in medicine (AIM) is 25 years old. Throughout this period, the field has attracted many of the best computer scientists, and their work represents a remarkable achievement. However, AIM has not been successful—if success is judged as making an impact on the practice of medicine. Much recent work in AIM has been focused inward, addressing problems that are at the crossroads of the parent disciplines of medicine and artificial intelligence. Now, AIM must move forward with the insights that it has gained and focus on finding solutions for problems at the heart of medical practice. The growing emphasis within medicine on evidence-based practice should provide the right environment for that change.

■ **JAMIA.** 1996;3:363–366.

Judged by the appearance of the first papers on the subject around 1971, the modern study of artificial intelligence in medicine (AIM) is 25 years old this year.[1] There are important lessons to be learned from this period, not only for researchers in AIM but also for the broader medical informatics community.

In reviewing the field in 1984, Clancey and Shortliffe provided the following definition:

> Medical artificial intelligence is primarily concerned with the construction of AI programs that perform diagnosis and make therapy recommendations. Unlike medical applications based on other programming methods, such as purely statistical and probabilistic methods, medical AI programs are based on symbolic models of disease entities and their relationship to patient factors and clinical manifestations.[2]

Affiliation of the author: Hewlett-Packard Research Laboratories, Bristol, England.

An earlier version of this paper was presented as the opening address at the IEE Colloquium on Artificial Intelligence in Medicine, Savoy Place, London, on February 19, 1996.

Correspondence and reprints: Enrico W. Coiera, MB, PhD, Hewlett-Packard Research Laboratories, Filton Road, Stoke-Gifford, Bristol BS12 6QZ, UK. e-mail: ewc@hplb.hpl.hp.com

Received for publication: 3/20/96; accepted for publication: 7/16/96.

Much has changed since then, and today this definition would be considered narrow in scope and vision. Today, the importance of diagnosis as a task requiring computer support in routine clinical situations receives much less emphasis.[3] So, despite the emphasis of much early research on understanding and supporting the clinical encounter, expert systems today are more likely to be used in laboratories and educational settings, for clinical surveillance, or in data-rich areas like intensive-care setting.[4]

Artificial intelligence, for its part, is less concerned now with the relative merits of reasoning under uncertainty, decision analysis, or symbolic or probabilistic reasoning. One is more likely to hear heated debates about software agents or the merits of building formal reasoning systems in which one can evolve simpler intelligence through interaction with the environment.

For its day, however, the vision captured in this definition of AIM was revolutionary. At that time, AI in medicine was a largely United States-based research community. Work originated out of a number of campuses, including MIT–Tufts,[5] Pittsburgh,[6] Stanford,[2] and Rutgers.[1] These researchers had a bold vision of the way AIM would revolutionize medicine and push forward the frontiers of technology. The field attracted many of the best computer scientists; by any measure, their output in the first decade of the field remains a remarkable achievement.

Throughout the 1980s, the output from what now had become a global research community was steady and substantial. Europe, in particular, had developed a significant research community that invigorated and broadened the scope of earlier research.[7] However, AIM was not successful—if success is judged as making an impact on the practice of medicine.

A period of soul-searching and redefinition of goals has followed. Sometimes AIM is characterized as an adolescent field,[8] almost as if to excuse the bold and youthful idealism of the early days and the apparent failure to deliver upon that idealism. People have thus begun to question whether the goals and spirit of AIM, as captured in the definition above, are still valid. At the heart of that debate sits a much wider and longer running debate on the relative merits of pure versus application-driven research. Both approaches have great merit, and both measure their success in different ways. Pure research strives to push forward the boundaries of knowledge, whereas application-focused work seeks to improve the processes that touch our daily lives. The fear grows that much of the recent work in AIM falls between these two great camps and, as a consequence, has a limited chance of making a substantial contribution to either.

Problems and Technologies

Artificial intelligence is a broad collection of technologies and goals. Its researchers work both to extend their understanding of the ways in which intelligent systems can be constructed and to apply that knowledge in the real world. Medicine, on the other hand, is a much older enterprise and is much clearer about its goals. Artificial intelligence in medicine is a hybrid field, formed out of the union of these two enterprises. United through AIM, these two communities can interact in three quite distinct ways.

First, the relationship can be technology driven. Medicine can provide AI researchers with a complex set of real-world problems with which to evolve their techniques. The outcome of good technology-driven research is the development of general purpose technologies that can be applied in many different domains. Those AIM researchers working in this way would judge their research a success if it were published in an AI journal. For example, much of the early work in computer-based diagnosis focused on medicine because it was such a good testing ground. Today however, AI researchers working on advanced diagnosis technologies have moved away from medical problems, and are more likely to be working on diagnosing faults in digital circuits, computer networks, or photocopiers.

Second, AIM can be problem driven. In this case, there are pressing medical problems needing solutions, and AI competes with other alternatives to provide those solutions. Success here is measured by the ability to solve real-world problems, and one would expect a researcher to be able to publish this type of work in mainstream clinical journals. Success would be measured by reductions in patient morbidity or mortality or by improvements in the efficiency of health care delivery. It should follow that otherwise technologically disinterested clinicians would be interested in the results of this work because it would be clinically relevant to them.

The third approach of AIM does not extend the boundaries of AI or solve the problems of medicine; instead, it is inward looking. Driven initially by the need to solve real medical problems, one may nevertheless need to solve technical issues that are not familiar to clinicians. At best, these problems can appear esoteric to clinicians; at worst, they can seem irrelevant. Equally, these technical problems might be inspired by AI but not be of major interest to it. Indeed, some of the most difficult work in AIM does belong right here.

For example, clinical outcomes need to be measured to improve medical practice. Consequently, ways of creating an electronic patient record need to be found so that clinical data can be pooled and analyzed. However, to achieve this, methods are needed to extract meaning from an often complex medical record. Consequently, there is a large effort, both in medical informatics and in AIM, to develop terminology coding schemes for electronic patient records. Separated by such a long-chain of reasoning from the original problem, this type of research usually appears in specialist journals devoted to medical informatics, and not routinely in medical or AI journals.

It would be fair to say that the dominant mode of AIM (and perhaps to a lesser extent Informatics as a whole) over the last 10 years has been in this third category. There are plenty of journals in which work is published for and by the informatics community. However, both AIM and informatics at present have a minimal impact on the fields that gave them birth. Have medical informatics and AI in medicine become so engrossed in their own problems that they have lost sight of the needs of medicine and AI? Or are they now valid research fields that can stand on their own without justifying their work within these other fields?

The answer has to be both. AIM and Informatics each has a unique set of technical problems to solve. They also, however, must judge what their labors achieve against their impact on the wider world. While medicine will always be there for researchers in AI to test their technologies, AIM is first and foremost a subdiscipline of medicine. Its achievements will thus ultimately only reach medical significance when they can demonstrate a positive impact on health care. The final arbiters of success must rest in phrases like "clinical outcomes" and "cost-effectiveness" and not in measures like "computational complexity."

Challenges

With an acceptance that AIM must adopt health care's problems as its own comes a redefinition of goals. For example, it is now generally recognized that, before AIM research can produce systems that have a significant impact, a substantial information infrastructure will have to be in place.[8] Consequently, some researchers have now moved from AIM to assist with the enormous task of developing an informatics infrastructure for health care.

At the same time that AIM is redefining its priorities, medicine itself is undergoing a quiet revolution known as "evidence-based medicine," and this may well transform the nature of clinical practice.[9] The mountain of research produced by medicine each week is now so great that the time lag between a treatment being proven effective and actually coming into routine practice is often measured in years. As a consequence, even the most diligent practitioners are unable to deliver care that represents the best available practice to their patients.

The problem lies in the fact that the mechanisms for transferring evidence into clinical practice are unable to keep up with the ever growing mountain of clinical trial data.[10] For example, the first trial to show that streptokinase was useful in treating myocardial infarction was published in 1959. Convincing evidence mounted in the early 1970s, and the first multitrial meta-analysis proving the drug's value was published in the early 1980s. However, formal advice that streptokinase was useful in the routine treatment of myocardial infarction only appeared in the late 1980s.[11] This was a full 13 years after a close examination of the published literature would have indicated the treatment's value.[12]

Many people in medicine see a move to universally available and codified guidelines for clinical practice as the solution to this problem. Not surprisingly, this poses both cultural and technical problems for health care. It should be enormously exciting to AIM researchers that medicine's technical problems are ones that AI may be able to help solve.

Developing large database of practice guidelines requires knowledge-based technologies to create and maintain them. Therefore, AIM's skills in knowledge acquisition and representation are needed to help develop methods that allow doctors to compare newly published data with existing guidelines and to update the knowledge base as appropriate. It may also be desirable to customize these knowledge bases to reflect local needs and conditions in different regions and countries. Work has already been underway in these areas for some time, but much remains to be done.[13-15] Ultimately, what is required is a way for practicing clinicians to access such guidelines quickly, incorporate them into their clinical practices, and then submit their own experiences back to the knowledge base to help improve it.

This raises one final challenge whose impact has yet perhaps to be fully appreciated—the rise of the Internet. The Internet is important for two reasons. First, it seems to be custom made to solve some of the inherent communications problems that are at the heart of creating a truly evidence-based medical practice.

Its second lesson is perhaps more subtle but may in the long run prove to be as important for AIM. Computer science, and AI in particular, is based on the power of formalizing knowledge representation and reasoning. Artificial intelligence in medicine is focused on developing ways of formalizing everything from diagnosis to the terminology medicine uses. Yet, the Internet's growth has probably been possible only because it formalizes so little. When it comes to defining how one should publish on the Internet, less is definitely more. Is there an equivalent level of semiformality for protocols and clinical guidelines? If so, this would allow some ability to manage and update guidelines while still making them flexible enough to use in different and sometimes unexpected ways.

In summary, through the rise of evidence-based medicine and the enormous challenges it poses, AIM and medical informatics have a new and important role to play. It may well be that in 25 years the question will not be, "What has artificial intelligence in medicine ever achieved?" but rather, "How could medicine have ever got here without it?"

References ■

1. Nordyke RA, Kulikowski CA, Kulikowski CW. A comparison of methods for the automated diagnosis of thyroid dysfunction. Comput Biomed Res. 1971;4:374–89.
2. Clancey WJ, Shortliffe EH. Readings in Medical Artificial Intelligence: The First Decade. Reading, MA: Addison-Wesley, 1984.

3. Durinck J, Coiera E, Baud R, et al: The Role of Knowledge Based Systems in Clinical Practice. In Barahona P, Christensen JP (eds). Knowledge and Decisions in Health Telematics—The Next Decade. Amsterdam: IOS Press, 1994; 199–203.
4. Coiera EW: Artificial Intelligence Systems in Routine Clinical Use. http://www-uk.hpl.hp.com/people/ewc/list.html
5. Szolovits P. *Artificial Intelligence in Medicine.* AAAS Selected Symposia Series., Boulder, CO: Westview Press, 1982.
6. Miller RA, Pople HE Jr, Myers JD. INTERNIST-1: an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med. 1982;307:468.
7. Stefanelli M. European research efforts in medical knowledge-based systems. Artif Intell Med. 1993;5:107–24.
8. Shortliffe EH. The adolescence of AI in medicine: will the field come of age in the '90s? Artif Intell Med. 1993;5:93–106.
9. Mulrow CD. Rationale for systematic reviews. BMJ. 1994; 309:597–9.
10. Wyatt J. Use and sources of medical knowledge. Lancet. 1991;338:1368–73.
11. Antman E, Lau J, Kupelnick B, Mosteller F, Chalmers T. A comparison of the results of meta-analysis of randomized controlled trials and recommendations of clinical experts. JAMA. 1992;268:240–81.
12. Heathfield H, Wyatt J. Medical informatics: Hiding our light under a Bushel, or The Emperor's New Clothes? Methods Inf Med. 1993;32:181–2.
13. Morris A, Wallace C, Menlove R, et al. A randomized clinical trial of pressure-controlled inverse ratio ventilation and extracorporeal CO2 removal from ARDS. Am J Respir Crit Care Med. 1994;149(2):295–305.
14. Musen MA, Tu SW, Das AK, Shahar Y. A component-based architecture for automation of protocol-directed therapy. Proceedings of the Fifth AIME Conference on Artificial Intelligence in Medicine—Europe. Berlin: Springer, 1995;3–13.
15. Renaud-Salis JL. Distributed clinical management-information systems: An enabling technology for future health care programmes. In Barahona P, Christensen JP (eds). Knowledge and Decisions in Health Telematics—The Next Decade. Amsterdam: IOS Press, 1994;139–46.