The melding of artificial intelligence and medicine

"One of the lessons learned in knowledge engineering is just how complicated is human knowledge"

Artificial intelligence in medicine — computer programs that attempt to act as medical experts or consultants — is passing through the laboratory door and is ready to begin appearing in the clinics.

Doctors' offices are likely to be next.

Since Stanford is one of the world's leading centers for AI (artificial intelligence) research, a lot of those laboratory doors and clinics are here. And some of the first patients whose treatment is influenced by these programs also are here.

At the Medical Center's outpatient oncology clinic patients with Hodgkin's disease, non-Hodgkin's lymphomas, and breast cancer, have been treated with the help of a program called ONCOCIN. The program keeps track of the patient's progress, advises on laboratory tests, and recommends drug therapies. Since almost no two cases of these diseases are exactly alike, and since many of the therapeutic drugs are toxic, physicians in the clinic, even those reluctant to use computers, generally accept the help.

The researchers are now adapting the program so it will run on the desk-top computers that doctors' offices can afford.

A program called PUFF, derived from another Stanford AI program, has been actually interpreting physiological tests of lung patients at the Pacific Medical Center in San Francisco. A program to help physicians appraise professional journal reports is under development.

Stanford, of course, is not the only place where researchers are trying to use AI tools to help physicians and medical researchers. The national interest in biomedical applications of AI became so intense in the early 1970s that the National Institutes of Health several years ago agreed to fund a national computer resource, a large machine that many researchers from around the country could access. They put it on the ground floor of the Stanford Medical Center, and it is known as SUMEX-AIM (Stanford University Medical Experimental Computer-Artificial Intelligence in Medicine).

SUMEX-AIM is a collection of scientific workstations tied to a Digital Equipment Corp. 2060 and VAX computers. Three to four hundred researchers around the country are plugged into the SUMEX computers.

The leading edge of this research at Stanford is at the Knowledge Systems Laboratory (KSL), an interdisciplinary set of projects tying the Computer Science Department to the School of Medicine.

More than 100 researchers, including five principal investigators, and scores of graduate students work in the five sublabs of the KSL. In addition, researchers in other fields, such as Oleg Jardetzky, of the Magnetic Resonance Laboratory, and Charles Yanofsky in biology, are working on specialized programs.

From the beginning, a goal of AI researchers has been to encapsulate the knowledge and problem-solving skills of experts into a computer application, so-called expert systems. It would be useful, the researchers believed, to try to put the reasoning and knowledge of human subject experts into the computer so non-experts could draw upon their skills.

The first expert system was produced at Stanford in the late 1960s. Called DENDRAL, the program was developed by Nobel laureate Joshua Lederberg, Bruce Buchanan and Edward Feigenbaum, professors of computer science. It attempted to capture Lederberg's expertise in analyzing organic compounds from mass spectroscopy.

One of the most successful AI programs ever, DENDRAL can be found in many organic chemistry labs, issued under a Stanford license.

DENDRAL is based on a core concept in artificial intelligence, the use of *heuristics*, sometimes known as the art of good guessing. Heuristics are similar to the mental processes of a chess player. The chess expert ignores the almost-infinite number of possible moves in a game, concentrating only on those relevant to the particular situation.

DENDRAL cannot consider every possible molecule in doing its analysis. Using heuristics, however, the program considers only those likely to be the answer.

The next logical step was a program for clinical medicine, designed as part of a dissertation by Edward Shortliffe, now associate professor of medicine and computer science. Shortliffe is an M.D. with a passion for computers. Along with Stanley Cohen, Bruce Buchanan, and Stanton Axline, Shortliffe developed MYCIN.

MYCIN, which uses heuristics, gives advice on the use of antibiotics for the treatment of infectious diseases.

Joshua Lederberg

Edward Feigenbaum

The method used to put the expertise of Cohen and Axline into the computer shows how these programs are developed — and how difficult it can be to build them. The process is called "knowledge engineering," and it turns out to be one of the most complicated parts of building AI programs.

"We sat around, went over patient cases, and tried to understand how Axline and Cohen would decide how to treat those cases," explained Shortliffe, who was a second-year medical student at the time. "We'd stop them — those of us who knew only a little medicine and were more computer scientists — interrupt and ask, 'Well, why do you say that?'

"We met once a week for an hour and a half. Axline would bring in an interesting case, a chart, and Cohen would ask him questions. Axline would answer them from the chart. Cohen would be problem-solving, trying to figure out what he'd do for that case, and we would try to understand why Cohen was asking the questions he was.

"We'd write down the rules that he told us, that Axline would help refine with him, and in the interim week, I'd put those rules into this developing and emerging computer system," Shortliffe said.

"And then we'd all have a good laugh the following week when I showed them how the computer had tried to handle the same case. What you did (was) discover the great simplifications they made in explaining the rules from the previous week, where they'd go wrong if you tried to run it on any kind of different case."

And, "every once in a while," Shortliffe said, "some issues would arise which caused a major change in the underlying structure of this developing program. Suddenly it wasn't just a matter of writing the rules; you had to make major program changes."

The program, like most AI programs, is written in a difficult, non-numeric computer language called LISP; the data and the commands are so integrated that by altering one, you may automatically alter the other.

One of the lessons learned in knowledge engineering is just how complicated is human knowledge. The computer, for instance, was quite capable of asking whether a patient was pregnant, even if the patient was male. We know only females get pregnant, but the programmers had to remember to add that little bit of commonsense knowledge to an intrinsically ignorant machine.

Nonetheless, MYCIN worked. More important, the program could be exported to other fields. The logical part of the program, what came to be called the "inference engine," seemed to be relatively universal. By extracting it and simply plugging in new data from a different domain, the MYCIN program could work for other fields just as it did for infectious diseases. The inference engine was called Essential MYCIN or EMYCIN. PUFF at the Pacific Medical Center is based on EMYCIN.

MYCIN also led to ONCOCIN, the cancer treatment program under development by Shortliffe and Lawrence Fagan.

ONCOCIN monitors the patient's condition by asking the physician questions about the patient, the treatment given so far, and the results of tests. The program manages the patient's drug therapy and is useful because of the complexity of cancer chemotherapy. The computer also has the ability to retain in its memory far more treatment details than an oncologist can remember. It is also constantly being updated with the newest research results.

If, for instance, a test shows that a patient's white blood cell count is decreasing, ONCOCIN may suggest ways of changing drug treatment. As with all such programs, the physician can accept or reject the advice — the ultimate responsibility remains with the doctor.

(That is not a minor issue. Besides the natural reluctance of a physician to relinquish control, particularly to a machine, the question arises: what happens if the machine is wrong? Who gets sued for malpractice? These are challenging legal issues not yet tested in the courts.)

ONCOCIN has a secondary program that explains the basis for its decision. If the physician asks why a recommendation is made, ONCOCIN reviews its reasoning and documents the information upon which the advice was given. The physician then has a better idea how much weight to give the suggestion — and, in some cases, may learn something he or she did not know, or remember something forgotten.

ONCOCIN is built on a series of IF-THEN statements; *if* a particular situation applies, *then* the computer should conclude something specific or recommend a particular course of action. However, "although ONCOCIN may appear to be very literal. . .it is important to remember that we don't tell the computer what to recommend for every conceivable situation," said Larry Fagan, project director and senior research associate.

"Instead, we supply it with knowledge and instruction on how to put bits of knowledge together. The computer then integrates incoming information on its own." In very large programs such as ONCOCIN, the result is often unexpected.

How well do these machines do? In a test matching ONCOCIN with human experts, the program "shows excellent performance of the computer relative to physicians treating cancer patients at Stanford," Fagan said. (The study was described in the December 1985 issue of *The Annals of Internal Medicine*.)

Another large expert system, CADUCEUS, written at the University of Pittsburgh, was tested using cases from *The New England Journal of Medicine*. Its performance was then compared to that of a group of physicians. Over a wide range of diagnoses, the machine was found to be more accurate than an average physician, roughly comparable to the teams of physicians who cared for the patients, and almost as good as expert physicians asked to review the cases in retrospect.

Although CADUCEUS can handle 600 different diagnoses, the scientists who programmed it, like those in all aspects of AI in medicine, do not believe the program can ever replace the physician.

Among other things, says Pittsburgh's Jack Myers, CADUCEUS lacks imagination. It cannot cope with a disease it is unprogrammed to diagnose, whereas the bright human physician will recognize something new. All the programs, including ONCOCIN, lack imagination.

ONCOCIN has been used experimentally in the Stanford oncology clinic since 1981, but the large computer prototype was recently removed in anticipation of the new version that will run on small machines.

The new interface with the physician will resemble the *Apple Macintosh* with graphics and "windows" (sections of the screen that show specific functions). The physician will move a mouse to control the action on the screen.

The screen will look just like the kind of paper chart the physician is familiar with — a deliberate design feature.

"It can do everything the physician's paper record can do, but is also able to do things that only computers can do — for example, using continuous forms on the screen that can be lengthened as necessary," Fagan says.

Edward Shortliffe

A separate research program is being developed to use more general strategies, Fagan said. This will enable the physician to cope with more complex cases. "For the simple case it comes out with a simple answer; for the complex case it comes out with more general or fuzzier answers.

"This corresponds to two cases. One is the regular case which the physicians generally handle themselves," Fagan said. "The other corresponds to the place where they would call in the expert for specialized advice."

Having ONCOCIN on a large research computer is of little practical value to a local physician; the goal is to put ONCOCIN on a workstation, a microcomputer-sized machine costing between \$10,000 and \$20,000, which puts the machinery within reach of many doctors' offices.

The Medical Computer Science Group has proposed distributing ONCOCIN experimentally in collaboration with the Northern California Oncology Group so the programs can be tested on workstations away from Stanford.

ONCOCIN is perhaps the largest of the medical AI projects going on at Stanford. However, there are several other related projects.

One program, called PATHFINDER, is an attempt to advise pathologists on the proper classification of lymph node abnormalities viewed under the microscope. This program is being built in collaboration with the University of Southern California.

REFEREE, a program being written by Bruce Buchanan of KSL and William Brown and Daniel Feldman of the Medical Center, is an attempt to try to build an adviser for interpreting papers from the medical literature. REFEREE will give the physician a feel for the reliability of the data in the reported study.

PROTEAN, developed by Buchanan and Barbara Hayes-Roth of KSL, along with about 10 students and nuclear magnetic resonance expert Oleg Jardetzky, will help researchers determine the structures of protein molecules.

In the meantime, SUMEX has been changing. A number of research projects from around the country have chosen to move off the large DEC machines at the Medical Center and use workstations or large micro computers.

SUMEX is now part of the Symbolics Systems Research Group, the sublab in the KSL that is responsible for the computing "environment" for the laboratory

Interestingly, although the medical programs, lik all the other AI programs at Stanford, frequently hav practical applications, the researchers are driven b the means, not the end. Trying to unravel Oleg Ja detzky's world in nuclear magnetic resonance ma prove helpful to Jardetzky, but it is more helpful to th AI researchers, who learn by doing.

"It's a fine line, because it looks like we're building a application for Jardetzky. That we hope will be a re sult, but it's not anything whose success we coul guarantee, certainly not when we started. It's true every project we do," Buchanan says.

On the other hand, of course, it would be nice if th programs work, he adds. That helps get enthusias collaborators.

All this could lead to a scenario out of science fictio

A doctor in the Sierras has a patient whose disease cannot handle. He needs help. The help comes fro his office computer. Housed in the metal box, on th *silicon chips, behind the video screen is an inanima* expert.

The computer asks questions. The physici answers them at the keyboard. The computer del into its vast data bank, makes the judgment worthy an expert, and prints out its best advice.

Ed Feigenbaum, the co-author of DENDRAL, b lieves *that* will begin the real computer revolutio

—Joel Shur