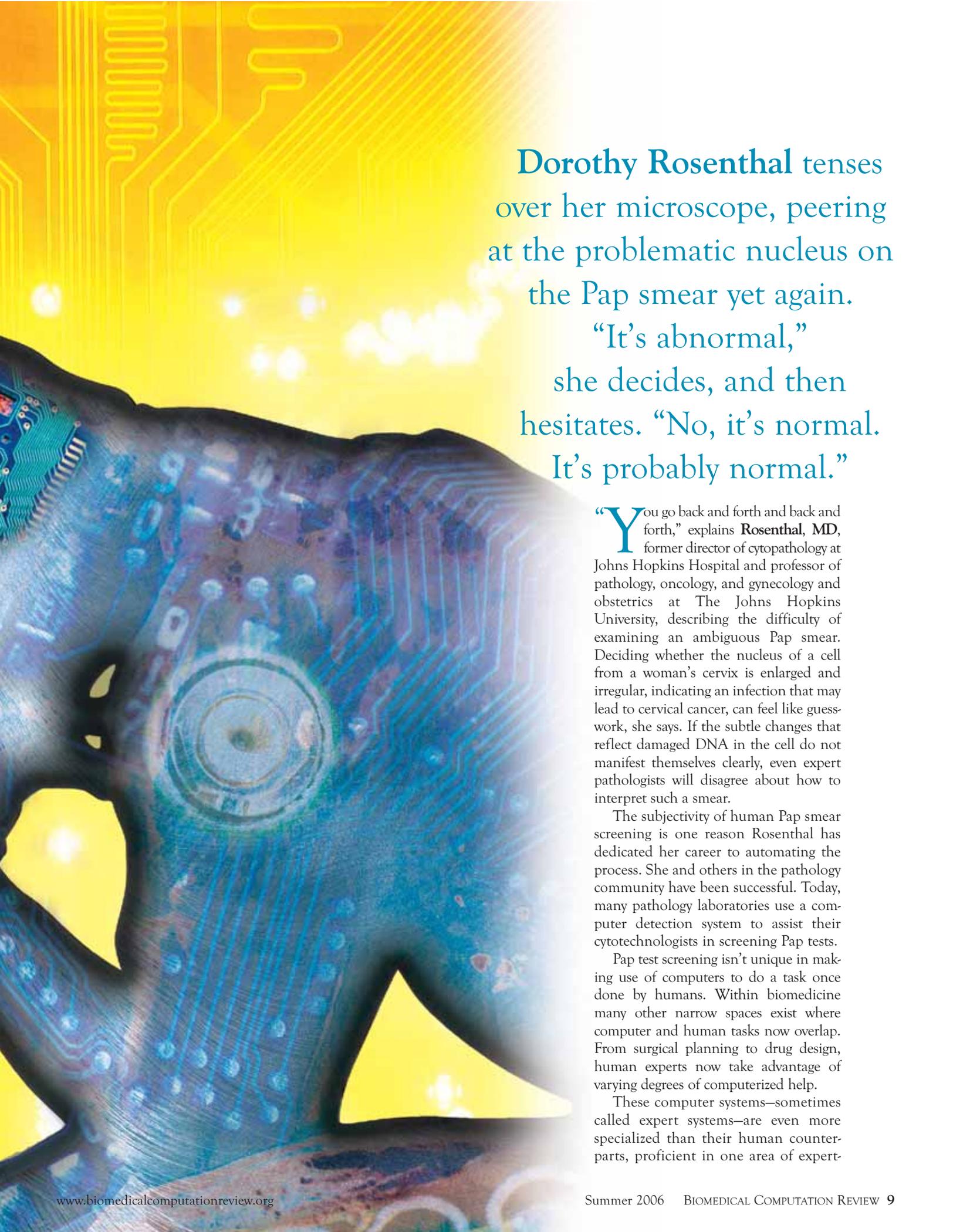


Biomedical
expertise
meets
computer
automation

BY LOUISA DALTON

Human vs Machine





Dorothy Rosenthal tenses over her microscope, peering at the problematic nucleus on the Pap smear yet again.

“It’s abnormal,” she decides, and then hesitates. “No, it’s normal. It’s probably normal.”

“You go back and forth and back and forth,” explains **Rosenthal, MD**, former director of cytopathology at Johns Hopkins Hospital and professor of pathology, oncology, and gynecology and obstetrics at The Johns Hopkins University, describing the difficulty of examining an ambiguous Pap smear. Deciding whether the nucleus of a cell from a woman’s cervix is enlarged and irregular, indicating an infection that may lead to cervical cancer, can feel like guesswork, she says. If the subtle changes that reflect damaged DNA in the cell do not manifest themselves clearly, even expert pathologists will disagree about how to interpret such a smear.

The subjectivity of human Pap smear screening is one reason Rosenthal has dedicated her career to automating the process. She and others in the pathology community have been successful. Today, many pathology laboratories use a computer detection system to assist their cytotechnologists in screening Pap tests.

Pap test screening isn’t unique in making use of computers to do a task once done by humans. Within biomedicine many other narrow spaces exist where computer and human tasks now overlap. From surgical planning to drug design, human experts now take advantage of varying degrees of computerized help.

These computer systems—sometimes called expert systems—are even more specialized than their human counterparts, proficient in one area of expert-

ise, at sea in all others. The IBM super-computer Deep Blue, which played and beat chess grandmaster Garry Kasparov in 1997, is such a system.

Some have wondered whether Deep Blue launched a computer revolution that would extend into all spheres. Are Pap test screening and chess-playing just the first of many arenas in which computers will one day outperform humans?

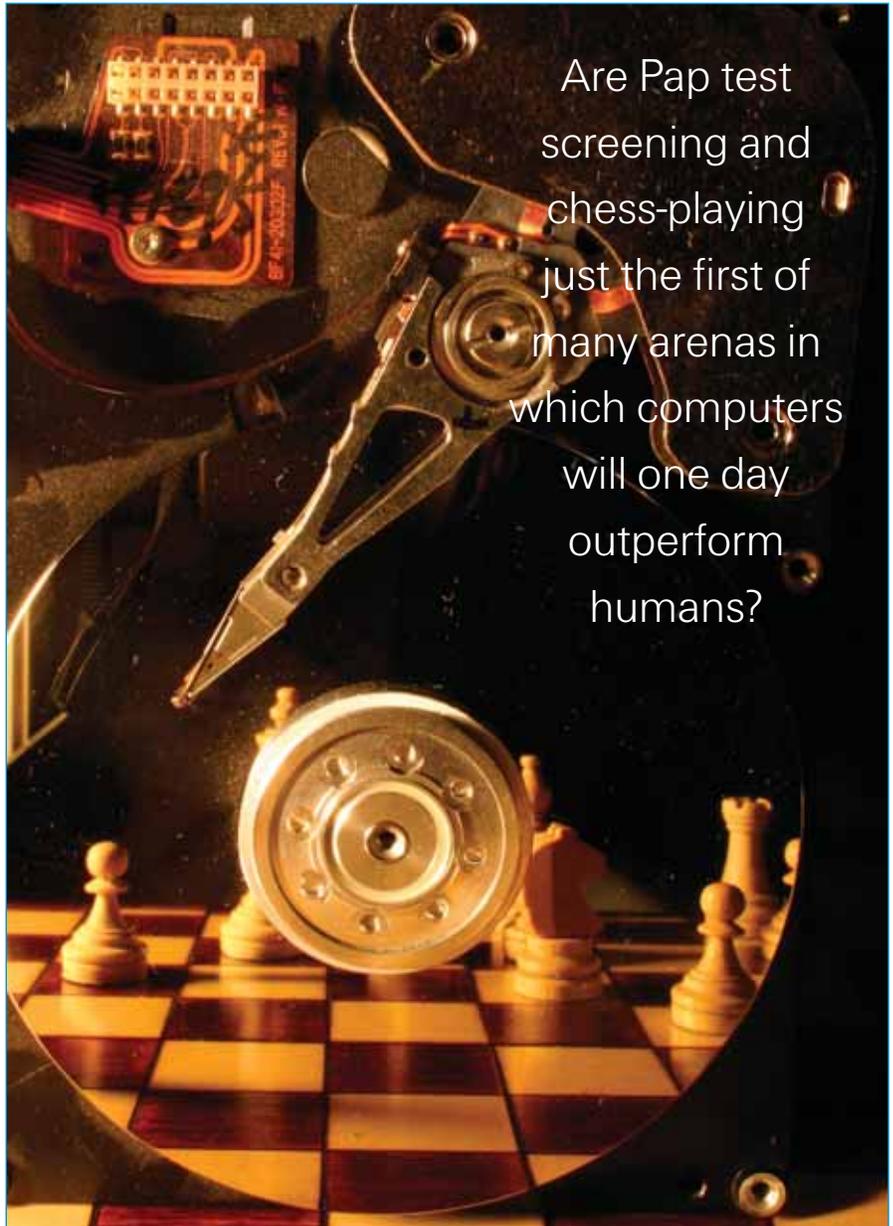
If so, it's a slow-moving revolution, at least in biomedicine. Although computers now rival, even surpass, human performance in some biomedical specialties, the challenges to widespread acceptance and use are still great.

There's the often sticky issue of changing a human's routine so that a computer can help. What the computer provides has to seem worth the trouble. Plus, people desire to stay sharp at their own specialty and want to safeguard against machine error. As a result, introducing computer automation in biomedicine has been extremely challenging on many levels.

Nevertheless, three examples—clinical diagnosis, image interpretation and protein structure prediction—illustrate the promise of developing computerized expertise. Computers can complement and augment native human capabilities and in some cases, even replace humans at the most mundane and repetitive tasks freeing scientists and doctors to pursue more interesting work.

MEDICAL DIAGNOSIS

Next to every bed at LDS Hospital in Salt Lake City is a computer terminal that gathers patient statistics: blood pressure, medications, ventilator activity and other key bits of information. The



Are Pap test screening and chess-playing just the first of many arenas in which computers will one day outperform humans?

hospital laboratory, the front desk, radiology, and physicians themselves.

Every time new data enter the system, the computer reevaluates patient status and decides, for example, whether or not to alert a doctor or recommend a medication adjustment. Physicians also use

chair of the medical informatics department at the University of Utah, says it has been working smoothly for years. The HELP system started operating in 1967 and is one of the pioneers of hospital decision support. In some of its specialized functions, it “provides more

Although computers that could perform a medical diagnosis and recommend treatment were among the earliest expert systems, the medical community has far from embraced them.

data are collected and managed by a hospital-wide information and decision support system called HELP—Health Evaluation through Logical Processing. The system also collects data from the

the system interactively for help with diagnosis, data interpretation, patient management, and clinical protocols.

Reed Gardner, PhD, one of the designers of the system, and former

consistent, uniform care to people than physicians do,” he says. Yet Gardner thinks he could count on one hand the number of other hospitals with a similar system. “It is really not as widespread as



The computer room at LDS hospital in 1965 housed the working hardware and backup hardware for its hospital information system, HELP. HELP's creators include, from left, Homer R. Warner, MD, PhD; T. Allan Pryor, PhD; and Reed M. Gardner, PhD. Courtesy of LDS Hospital, Salt Lake City, Utah.

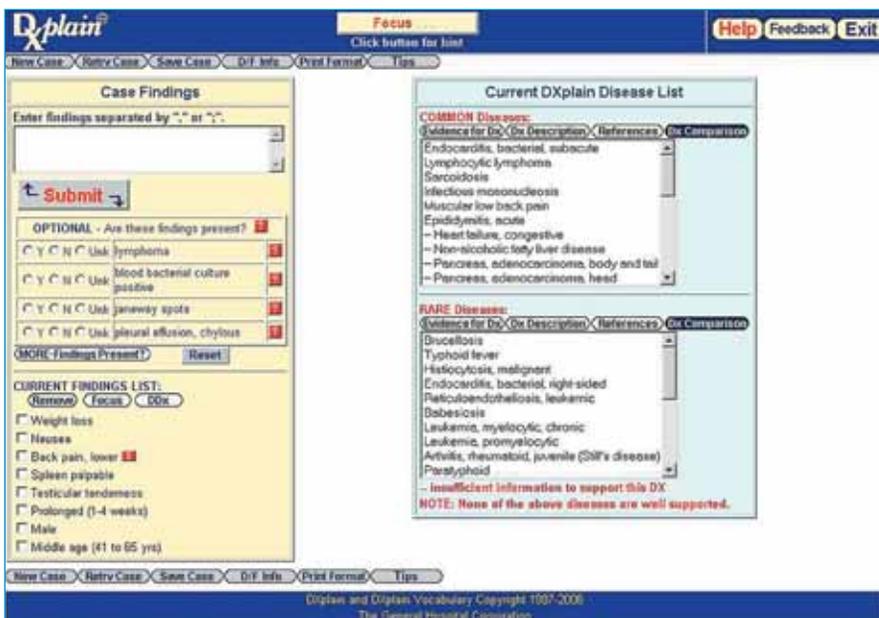
I would imagine 40 years later," Gardner says. Although computers that could perform a medical diagnosis and recommend treatment were among the earliest expert systems, the medical community has far from embraced them.

One of the sticking points relates to those data-gathering bed monitors at LDS Hospital. Gardner went to a great deal of trouble to create the terminals—even building many of the first ones with his own hands. He knew that if the

bedside computer wasn't automatically collecting data and sending the information to HELP, a doctor or nurse would have to type it in. That would have required a change in workflow. When Gardner first tested HELP, he actually hired computer technicians to do the data input for the doctors.

"Physicians are intensely practical," says Octo Barnett, MD, a developer of DXplain, a decision support system primarily for diagnosis that has been operating at Massachusetts General Hospital for more than 18 years. "They won't do something that takes a lot of time and effort to do and doesn't have a lot of payoff for them."

On top of the inconvenience of data entry, the perceived benefit of using



DXplain, a decision support system primarily for diagnosis, was developed at the Massachusetts General Hospital in 1986 and has been available internationally over the Internet for the past 10 years. Physicians can submit symptoms (left) and the system suggests possible diagnoses (right). Courtesy of Octo Barnett, MD, Massachusetts General Hospital.

“We feel that [the machine] helps us look harder at the cases that are most likely to be abnormal,” Nance says of the Pap test screening system used by Rex Healthcare.

such a system is low, according to **Eta Berner, EdD**, professor in the health informatics program at the University of Alabama at Birmingham. Computerized diagnosis, she says, solves a problem that some don't think needs a solution. Berner studies diagnostic errors and believes that many such errors go undetected by physicians. If your doctor doesn't get a diagnosis right the first time, she says, either you will go to another doctor, to the hospital, or back to your doctor who will try something else until you get sick enough that the correct diagnosis is obvious. None of those scenarios conveys to the physician that an error was made.

However, there is a growing awareness that medical errors, including misdiagnosis, are indeed a problem. In 1999, the Institute of Medicine (IOM) of the National Academies of Medicine released a startling report stating that medical errors cause an estimated 44,000 to 98,000 deaths a year in U.S. hospitals. The types of errors include misdiagnosis, incorrect drug dosing, equipment failure, infections, blood transfusion related injuries, and misinterpretation of a medical order.

Since then, the Institute of Medicine

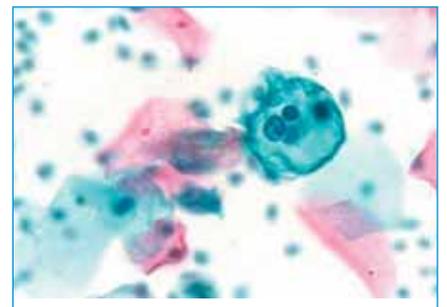
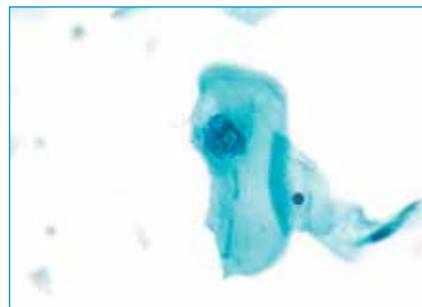
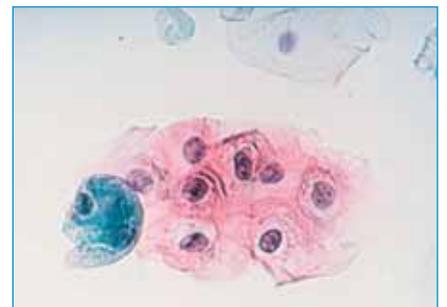
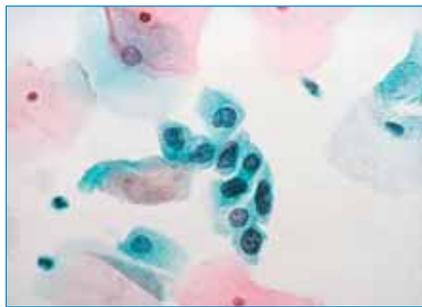
has called for computerized healthcare tools that effectively capture patient information and offer decision and diagnosis support aids.

Hospitals, nursing homes, and doctor's offices have begun to respond. They have preferred systems that, like HELP, primarily alert, remind, inform, and suggest, not just diagnose. Such activities fit into clinical practice better partially because they augment the staff's efforts, rather than replacing them. Berner calls these systems the low-hanging fruit: recommending the most cost-effective antibiotic, advising the pharmacist on drug-drug interactions, double-checking blood types before a transfusion, or carefully guiding and monitoring as a patient is weaned from a ventilator.

The doctors are generally positive if the system works well, Gardner adds, but they don't like it if one of the functions over-alerts, a server is down, or data doesn't get properly entered and lab results get delayed.

INTERPRETING MEDICAL IMAGES

When pathologist **Keith Nance, MD**, and his coworkers at Rex Healthcare were



Liquid-based Pap tests can vary from slides that clearly indicate a precancerous state (many abnormal cells with enlarged nuclei and irregular contours), as shown at top, to slides that are much more ambiguous because they contain only one or two atypical, abnormal-looking cells, as shown at bottom. Courtesy of Dorothy Rosenthal, Johns Hopkins Hospital.

told that the Pap test screening system they just bought could detect abnormal cervical cells better than humans could, it wasn't that they didn't believe it. They simply wanted to make sure for themselves. So for a few years, they double-checked everything the computer analyzed. Out of more than 100,000 cases, the machine missed only one case of the type of abnormal cell called high-grade dysplasia, less than 0.00001 percent of occurrences.

"Now that's good," Nance says, "because the human miss-rate is considered to be five to 10 percent." The computer wasn't quite as good at picking up low-grade dysplasia; it missed about three percent of them. Still, humans miss roughly five percent, Nance says. "Basically, we proved that the machine is better than humans."



That's why the machine that Rex Healthcare purchased, the FocalPoint slide profiler sold by TriPath Imaging, is FDA-approved to independently sign off 25 percent of the slides it sees. It dubs them as requiring "no further review" by human or machine. In addition, it ranks the remaining 75 percent from most likely to be abnormal down to least likely. Nance and his coworkers are pleased with the machine. "We feel that it helps us look harder at the cases that are most likely to be abnormal," he says, and rescreen the cases that really should be rescreened.

Pap test screening is a bright success story of computer assistance. Computer automation of the task is well accepted and cost-effective. It tackles a task for which there aren't enough people to do the work. And because routine Pap test screening "consists of long, tedious intervals between interesting cases," says Rosenthal, most humans gladly welcome help.

Effective, automated Pap test screening is now a reality, but getting to this point wasn't easy. It has taken patience and consistent funding. Back in 1979, the National Cancer Institute sent out a call for proposals to develop automated Pap smear screening systems. At that time, however, "the computers weren't capable of doing the kind of number-crunching we needed them to do," Rosenthal says. In 1987, an exposé in *The Wall Street Journal* about widespread

poor practices for Pap smear screening led to a public outcry and an even greater interest in automation. The U.S. government funded many groups from the late 1970s to the late 1980s. From 1990 to now, private money has developed the devices, Rosenthal says.

Despite the long incubation period and the cautious mistrust displayed by groups like Nance's, computer automation of cervical cancer screening survived. Machines offered by both TriPath Imaging and Cytoc Corporation now increase the productivity of the cytotechnologists and pathologists.

A radiologist's search for telltale signs of breast cancer (such as breast calcifications, tumors, or other lesions) on a film from an x-ray mammogram shares some

miss something obvious. The problem is very appropriate for a computer because computers search "pixel by pixel, area by area, without getting phone calls, without getting tired," says **Maryellen Giger, PhD**, professor of radiology at the University of Chicago.

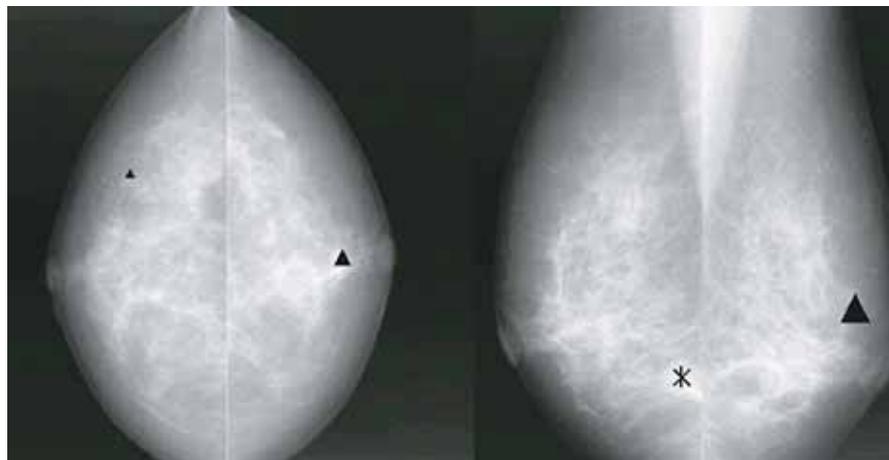
Giger develops computer algorithms and software to aid the radiologist. For more than 20 years, she has worked on computer-aided detection and diagnosis (CAD) of radiological images.

The first computer system that could search a mammogram for breast cancer was FDA-approved in 1998, and Giger estimates that a computer reads approximately a quarter of the screening mammograms performed in the United States nowadays. Unlike cervical cancer screening, however, the physician always has the first look at a mammogram. After the unaided radiologist searches carefully for cancer on a film,

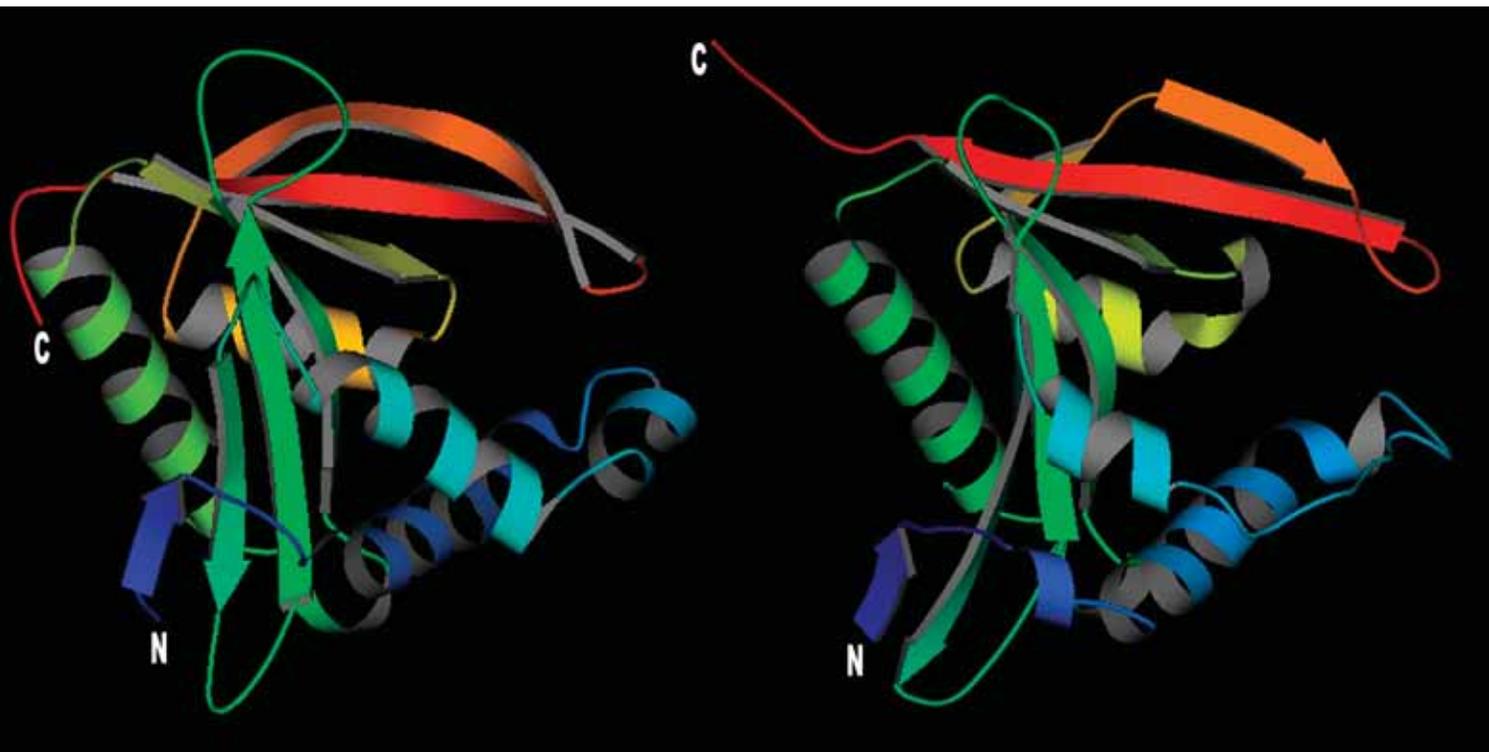
"If you let the computer do it first, there is the possibility that the radiologist gets lazy," says Sandy Napel. For the present, "it is thought best to put the radiologist in competition with the machine."

challenges with the pathologist's search for abnormal nuclei on Pap tests. With such long intervals between abnormal cases, it is easy for a human to get distracted, give up too quickly, or simply

the computer outputs the analysis of the digitized film, designating possible cancers. The radiologist then rechecks the image and either accepts or rejects the computer's suggestions. Most studies show



After reviewing a mammogram, a radiologist may activate a system such as R2 Technology's ImageChecker, which places asterisks at possible cancer masses and triangles at calcifications indicative of cancer. The larger the mark, the greater the likelihood of cancer. Courtesy of R2 Technology.



A predicted structure of an acetyltransferase generated by an automated server in Daniel Fischer's laboratory was among the best of the entries in CAFASP. The predicted structure (right) differs from the experimental structure (left) by just a few angstroms. Courtesy of Daniel Fischer, University at Buffalo (The State University of New York) and Ben Gurion University in Israel.

“Biologists don’t want to write to the winner of CASP and ask him to spend weeks modeling their particular protein,” Fischer says. “They want to go to the winner of CAFASP on the Internet and push a button.”

that by using this system, radiologists potentially catch more cancers, catch them earlier, and catch them in younger women.

The computer’s second-reader status in mammography actually adds some time to the radiologist’s review. And even though the computer catches calcifications better than most radiologists, for the moment, radiologists have resisted the idea of allowing the computer the first shot. “If you let the computer do it first, there is the possibility that the radiologist gets lazy,” says **Sandy Napel, PhD**, professor of radiology at Stanford University School of Medicine. “The radiologist must take a very close look at the images,” he adds, and being told where to look might keep the radiologist from looking closely everywhere. Even with the current system, “there is concern that as radiologists get more comfortable with the technology and see how effective it is at finding lesions, they may press the second reader button sooner.” For the present, Napel says, “it is thought best to put the radiologist in competition with the machine.”

PROTEIN STRUCTURE PREDICTION

For years, the ongoing joke among computational biologists was that the

protein-folding problem had again been solved that year. The long-standing problem consists of predicting the final, balled-up form of a protein given only its linear, amino acid sequence.

The problem of predicting protein structure is now a bottleneck to progress. Plenty of amino acid sequence data are being generated by genome projects, but computers can’t yet use that information to predict protein 3D structure—a valuable piece of information for rational drug design. Although researchers can get to final protein structures experimentally with x-ray crystallography, trying to crystallize and determine structures for all proteins (even just the hundreds of thousands of human proteins) will simply take too long given current technologies. Many scientists believe that we need a computational solution.

During the 1980s, groups regularly developed models that worked well on one protein only to find that the model didn’t work for every protein. Finally, in 1993, a pair of researchers got frustrated enough to declare a competition. **John Moult, PhD**, at the Center for Advanced Research in Biotechnology at the University of Maryland, and **Krzysztof**

Fidelis, PhD, director of the Protein Structure Prediction Center at the University of California, Davis, set up the Critical Assessment of Techniques for Protein Structure Prediction (CASP): an open competition held every other year where prediction groups compare their strategies head-to-head on new proteins. The experimental structures are published at the end of each competition, clearly revealing which groups performed well and which did not.

Most submissions come from research groups that use a combination of modeling

They've gotten better—so much better that in the latest CASP/CAFASP competition, “only a handful of human predictors did better than the best of the servers” in one of the prediction categories, Fischer says. The difference between a prediction made by an automated server and that of a human expert who chooses which programs to run and improves the results manually is getting “smaller and smaller,” Fidelis says.

Still, no predictions, whether submitted by human or machine, yet reach the quality of an experimentally deter-

puter take over a task so that humans can work on other problems. Losing a human skill because of technological advances is something humankind has been doing for a long time: from the loss of hand spinning with the invention of the spinning jenny to the loss of slide rule proficiency with the invention of the calculator.

Yet some, even some of those same researchers who support computer automation in biomedicine, also worry about the loss of a human art. If the computers at the pathology laboratory

If the computers at the pathology laboratory were to break down for two days, asks Napel, would we have enough cytotechnologists and pathologists to screen all the Pap smears?

programs, prediction algorithms, human familiarity with protein families, and gut instinct to come up with their predictions.

Yet increasingly, predictions are also coming from computers alone—automated servers that, except for help in setting the initial parameters, receive no human input at all. **Daniel Fischer, PhD**, a professor of bioinformatics at the University at Buffalo and the Ben Gurion University in Israel set up a parallel competition with CASP solely for automated servers. CAFASP (Critical Assessment of Fully Automated Structure Prediction) runs at the same time as CASP and uses the same data, making it “not only a competition of who is the best server, but also a competition of humans versus machines.”

Initially, not everyone liked the idea of automated servers competing with the human predictors. Yet Fischer felt strongly that the automated servers needed their own place in the competition because automation has to be the ultimate goal of the field. “Biologists don’t want to write to the winner of CASP and ask him to spend weeks modeling their particular protein,” he says. “They want to go to the winner of CAFASP on the Internet and push a button.”

The first year that CAFASP ran alongside CASP, the server predictions were downright lousy, Fischer says.

mined structure. The best predictions of both humans and computers position the backbone at least 1 to 1.5 angstroms away from where it ought to be. That’s good enough for some tasks, Fischer says, such as predicting how a protein assembles in a complex, but not yet good enough to create a drug that will act on the protein. “We still hope that someone will figure out how to do the last bit,” Fischer says.

Eventually, it is inevitable that automated servers will take over the task of protein structure prediction, and CASP competitors will have worked themselves out of a job.

Fischer is not very nostalgic about it. He says that both computationalists and biologists will then be freed up to work on weightier issues. We don’t worry about letting a calculator compute cubed roots for us, he says. “Structure prediction itself is a wonderful problem, I love it. But it is not the big picture. The big picture is, do you know what the protein does? Do you know how to suggest a drug to interact with it? I’ll be very happy if no human ever does that again by hand, and researchers concentrate on the more interesting and challenging problems of the 21st century. Protein structure prediction is a 20th century problem.”

Many researchers echo, to some degree, Fischer’s willingness to let a com-

were to break down for two days, asks Napel, would we have enough cytotechnologists and pathologists to screen all the Pap smears? Will we become too dependent on computers? Napel’s concern is a common one, and one reason that the automation of biomedical tasks is a slow-moving trend.

Rosenthal says that Napel’s concern is valid; we probably would not have enough humans at hand to screen the Pap tests if the machines broke down. But the benefits outweigh the loss, she says. The cytotechnologists who used to do mass screening are focusing more on the interesting, abnormal cases, she adds. Novel molecular and genetic tests are being applied to cytology samples, including Pap tests. Salaries will go up as cytotechnologists learn more and their skills become more valuable. Though they lose a task, their skill improves elsewhere, and the field advances.

Chung-Jen Tan, PhD, senior manager of the Deep Blue development team at IBM, referred to a similar effect in the world of chess shortly after the 1997 match between Deep Blue and Kasparov. He pointed out that there was more to the victory than just a game of chess. “This will benefit everyone,” he said, “from the audience to school children, to businesses everywhere, even to Garry Kasparov.” □