

Can machine learning bring more humanity to health care?

September 20 2018, by Kris Newby



Palliative care physician Stephanie Harman, left, believes in helping people die on their own terms. A computer algorithm that informatics expert Nigam Shah, right, designed to help predict hospital patient deaths enables palliative care teams to provide better end-of-life counseling. Credit: Stanford University

Stephanie Harman, MD, a palliative care physician at Stanford Hospital,

has witnessed many people take their last breath, and she considers each passing a unique and sacred event.

She once saw a father look into the eyes of his adult daughter and say, "I love you so much," then die seconds later. A hospitalized man with an aggressive cancer worked diligently to settle his affairs, then after he finished, paused for a few seconds and whispered in her ear, "Why am I still here?" Another daughter gently took her father's hand and said, "Dad, I'll be OK. It's OK for you to go now." The father obediently closed his eyes and passed away.

One morning Harman, petite with shoulder-length black hair and wearing a freshly pressed white coat, was called to the bedside of a 79-year-old man who had entered the hospital during the night with pneumonia. He had heart disease, diabetes and emphysema. He was on oxygen and an IV drip.

With a compassionate smile, she asked if he was up for a conversation. She knew an end-of life discussion needed to happen soon, but the family was too overwhelmed with emotions to broach the topic.

Harman began by asking, "What are your hopes if you can't get better?"

The patient's wife had recently had a stroke and he didn't think she would be able to [care](#) for him at home. Yet he wanted to be with her during his last days, not in the hospital, even if that meant he might not live as long. He had no advance medical directive, a legal document that specifies who should make decisions for him if he became incapacitated. So, Harman, the primary care team and a palliative-care social worker spent hours helping him find a home hospice service that would support his medical needs and his own plan for the end of his life.

Harman's patient was fortunate; all too often [patients](#) die in an intensive

care unit with unfinished business and missed goodbyes.

Harman is a co-leader of a Stanford pilot program that aims to change that. Each morning she receives a priority report from an intelligent computer program that, every 24 hours, analyzes which patients under the care of the general medicine physicians would benefit from palliative care. The tool helps her spend more time with patients and less on record reviews and, most importantly, it leads to better endings.

This is one of many Stanford Medicine projects that combine [artificial intelligence](#) technologies with medical expertise to help doctors make faster, more informed and humane decisions. They hope it will help them spend less time in front of computer screens and more time doing what they love: caring for patients.

The data wrangler

Harman was first introduced to the concept of AI in palliative care when Nigam Shah, MBBS, Ph.D., an associate professor of biomedical informatics, attended a hospital quality meeting in early 2017.

"I've developed a computer algorithm that predicts the likelihood of patients dying within 12 months," declared Shah, with a boyish face, a bulldog-like demeanor and onyx-black eyes that seemed to take in everything and everyone in the room. "Would you find that useful?"

Harman blinked, then said, "Yes. Yes. Physicians are terrible at predicting death."

Shah began developing a mortality prediction tool to help palliative care professionals identify patients who might benefit from having end-of-life conversations well before a medical crisis strikes.

Shah grew up in a small town in Gujarat, India's westernmost state, in an upper-middle-class family. His father was a surgeon who felt duty-bound to perform pro bono procedures for the poor. His mother was a teacher and a school principal.

Shah planned to become an orthopedic surgeon and trained as a doctor, obtaining a bachelor of medicine and surgery degree from Baroda Medical College in Gujarat. But a family friend convinced him to pursue a Ph.D. in the United States first.

He landed at Penn State in 2000 and was so intrigued by the Human Genome Project, the mapping of the 3 billion nucleotide base pairs that make up human DNA, that he convinced his Ph.D. committee to let him work on bioinformatics, the relatively new science of analyzing complex biologic data. For his thesis, he wrote a clever artificial intelligence program that predicted yeast behavior, and one of his Ph.D. committee members suggested a way forward for him: "All the physicians who like artificial intelligence work at Stanford."

In 2005, Shah joined the biomedical informatics lab of professor Mark Musen, MD, Ph.D., at Stanford. The university had been applying artificial intelligence to health care problems since the 1980s, after setting up the legendary Stanford University Medical Experimental computer for Artificial Intelligence in Medicine, called the SUMEX-AIM.

In the late 1990s, Musen and his colleague Mary Goldstein, MD, developed ATHENA, one of the first intelligent decision-support systems for managing patients with chronic diseases, such as hypertension. It's still in use at the Veterans Affairs Palo Alto Health Care System.

Stanford is also where three pioneers in statistics—Bradley Efron, Ph.D.;

Trevor Hastie, Ph.D.; and Robert Tibshirani, Ph.D.—developed algorithms to analyze complex data sets, laying the foundation for today's machine learning and data mining.

Stumbling into this AI hotbed just as electronic health records systems were taking off was the "aha moment" for Shah, who thought, "What if the evidence that physicians needed was buried deep within the vast, messy electronic health databases, and AI could help pull it out?"

"In hindsight this sounds like a brilliantly planned career trajectory, but it's nowhere close. It was a uniquely Stanford fluke," said Shah.

Predicting mortality

Shah began thinking about mortality prediction while working with an advanced-illness management group at a nearby hospital. A cursory search of the medical literature confirmed his suspicion that physicians are woefully inaccurate at predicting how long terminally ill patients will live.

One of the best research studies on this topic asked 343 physicians to estimate the survival time frame of the patients they'd referred to hospice. Only 20 percent of the prognoses were accurate. What's more, the physicians overestimated survival times by a factor of five.

The lead author of the study, Nicholas Christakis, MD, Ph.D., a professor of sociology and medicine at Yale University, went on to explore the reasons behind this over-optimism in the book *Death Foretold: Prophecy and Prognosis in Medical Care*. He attributed this inaccuracy to "a complex set of professional, religious, moral and quasi-magical beliefs." Or, put more simply, the physician's innate desire to never give up fighting for their patients' lives.

To bring some objectivity to prediction, some doctors use palliative scorecards that assign weighted mortality scores to a patient's observable symptoms. One system rates walking ability, level of self-care, food and fluid intake, and state of consciousness. Another assesses weight loss, breathing problems and white blood cell counts. Yet another calculates a risk based on food intake, swelling of tissues, delirium and breathing at rest. And for use in intensive care units, the Acute Physiology and Chronic Health Evaluation 2, or APACHE-2, assesses acute physiology, age and chronic health conditions.

Shah had issues with all of the scorecards. Some used data sets that were too small. Some used oversimplified assumptions. Others narrowly focused on specific diseases or populations. He wanted a tool to predict the probability of death of every patient admitted to the hospital every day, by comparing their medical record to the millions of past patients of the hospital. So, he opened his artificial intelligence toolbox and settled on supervised deep-learning approaches to determine the most important predictors of mortality.

Deep learning is a technique that allows a software algorithm to automatically discover important factors from vast arrays of raw data. When it's "supervised," the algorithm is allowed to analyze variables associated with known outcomes so that it can learn from the past and apply its findings to future situations in a repeatable way.

In developing the tool, Shah first formulated a problem statement to guide his algorithm: "Given a patient and a date, predict the mortality of that patient within three to 12 months from that date, using electronic health record data of that patient from the prior year."

Then he had it search and learn from the anonymized medical records of the millions of patients who entered Stanford hospitals between 2010 and 2016, comparing past mortality factors with those of a newly

admitted patient. For Shah's tool, the target outcome was a mortality prediction, and the variables included medical record entries such as an insurance code for a specific disease, a drug prescription or the pattern of visits. Here's how the system works:

Patient X is admitted at 9 p.m. At midnight, the algorithm looks at X's medical record for the past year and pulls out such features as age, gender, race, ethnicity, number of hospital admissions, disease classification codes, and billing and prescription codes. It aggregates those in groups over the past 30 days, 90 days, 180 days and beyond. The algorithm then compares Patient X's features with the combinations of features seen in millions of past patients and their subsequent outcomes. Finally, the software model calculates a probability of Patient X dying in the next three to 12 months.

The first set of results from Shah's algorithm were pretty good. Flags for high mortality risk included diagnosis codes of certain cancers and MRI and CAT scans, and multiple hospital admissions in a year. But there were obvious errors. A patient was put on the near-death list because an MRI scan was ordered under a brain tumor insurance code, even though the doctor later entered "No brain tumor" into the record.

But Shah didn't correct the inputs to the algorithm. "The algorithm needs to learn to handle such cases," he said, explaining that the algorithm would learn from its mistakes over time.

The value of palliative care

As Shah waited for his deep-learning algorithm to hone its prediction skills, Harman continued to struggle with the day-to-day challenges of a palliative care physician.

She became interested in the specialty during her first year of medical

school when her father-in-law was diagnosed with stage-4 lung cancer.

"The thing that stuck out in my mind was how he was able to die on his own terms," said Harman. Once it was clear that he wouldn't recover, he stopped treatment and visited his family cabin in Ontario one last time, then died at home with hospice care.

"He was the first patient I ever pronounced dead," said Harman.

She thought that everyone died with such dignity, with their wishes honored, but after graduation, she realized it was the exception, not the rule. Studies show that 80 percent of Americans want to spend their final days at home, but only 20 percent do. She went into palliative care to help others experience death according to their wishes and to live well until that time.

"It's terrible when you have to make decisions in crisis, because you may end up with medical care that doesn't match up with what matters to you, and you won't have time to think through the complex options."

Palliative care physicians are able to discuss death with kindness and clarity in a way that can make some doctors feel uneasy. Doctors are often fighting for a patient's life; a palliative care doctor is fighting for a patient's quality of life.

But there's a shortage of palliative care professionals in the United States. The National Palliative Care Registry estimates that less than half of the 7 percent to 8 percent of the admitted hospital patients who need palliative care actually receive it.

All of this factored into Harman's desire to work with Shah on an AI model that predicts the need for palliative care.

"Ideally with this AI model, we're identifying patients who are sicker than we realize," she said. "And it gives us an excuse to say, 'It'd be great if we could talk about advanced care planning.' Or, 'Have you had a discussion with your regular doctor about what matters most to you if and when you get sicker?' I think the twist is that we're using machine learning to add more to a patient's care without taking anything away."

The need for transparency

The tantalizing promise of being able to extract real-world clinical evidence faster and cheaper than the old ways motivates Shah to push his physician colleagues out of their comfort zone in embracing these new AI technologies.

"It bothers me that even in a well-studied field like cardiology, only about 19 percent of medical guidelines are based on good evidence," said Shah. "Much of it comes from trials that have focused on 55-year-old white males. For the rest of humanity, physicians make a best-faith effort, enter it in the [medical record](#), then never look back."

Robert Harrington, MD, professor and chair of medicine and an interventional cardiologist, believes that AI can help fix this, saying, "Clinical trials tell you about populations, not about that patient sitting in front of you. This is where machine learning comes in. It allows you to look at large volumes of aggregated records from the recent past and create models that can assist with predictions about that one individual."

The Achilles heel of today's AI tools, however, is that they're not that good at cause-and-effect reasoning. For example, an AI algorithm can't tell if a rooster's crowing makes the sun rise or the other way around. This is why having human experts involved in tool development is essential.

Case in point: When Stanford researchers first tested an AI tool for identifying cancerous moles, they were astounded at its accuracy. But when researchers analyzed the results, they identified a major flaw in the way they were training the algorithm: A large percentage of the cancerous mole photos had rulers in them. The algorithm drew the conclusion that rulers are a sign of cancer, not that physicians were more likely to use rulers to measure moles suspected of being cancerous. To correct this oversight, subsequent testing was done on photos without rulers in them.

The other risk with AI algorithms is that only clinicians with solid computer science know-how understand how they work, and this can lead to outcomes with unintentional biases or hidden agendas.

In the transportation industry, two news stories about algorithms with dark secrets buried in the code were described in a perspective piece that appeared March 15 in the *New England Journal of Medicine*: "A recent high-profile example is Uber's software tool Greyball, which was designed to predict which ride hailers might be undercover law-enforcement officers, thereby allowing the company to identify and circumvent local regulations. More complex deception might involve algorithms designed to cheat, such as Volkswagen's algorithm that allowed vehicles to pass emissions tests by reducing their nitrogen oxide emissions when they were being tested."

In health care, the stakes are even higher. Non-transparent "black box" algorithms could be used to deny care to certain classes of people, overprescribe certain high-profit drugs or overcharge insurance companies for procedures. Patients could be harmed.

This editorial and another in *JAMA* on Jan. 2 are part of a larger effort by Stanford researchers to address ethical issues to reduce the risk of these negative consequences. The authors include Shah; Harrington;

Danton Char, MD, assistant professor of anesthesiology, perioperative and pain medicine; Abraham Verghese, MD, professor of medicine; and David Magnus, Ph.D., director of the Stanford Center for Biomedical Ethics and professor of medicine and of biomedical ethics.

These experts warn that feeding biased data into an algorithm can lead to unintentional discrimination in the delivery of care. For example, the widely used Framingham Heart Study used data from predominately white populations to evaluate cardiovascular event risk, leading to flawed clinical recommendations for nonwhite populations.

"If we feed racially or socioeconomically biased data into our algorithms, the AI will learn those biases," said Char.

Humanity at the end of life

Harman now uses the second generation of Shah's palliative prediction tool. Each morning, it emails her a list of newly admitted hospital patients who have a 90 percent or higher probability of dying in three to 12 months. There are no names on the email or details about why they're on the list. It's up to Harman to review the medical records she receives and decide if those patients have palliative care needs. She's found the list to be helpful, and she sees how it can improve hospital care and enable her to spend more time with the most critical patients.

"Human physicians are way better at predicting death within a few days, but I'd bet on my model over a physician any day in predicting death three to 12 months out," Shah said.

The algorithm design and preliminary results of the first pilot study were published online in arXiv on Nov. 17, and another Bay Area health-care institution will be soon be piloting the algorithm.

"This isn't a population of patients with a single disease or more predictable courses of illness," said Harman. "The patient might have five or 10 different problems that are all interacting with one another—not just a stroke, but also cancer and emphysema, for example. With this model, it looks over a longer time frame, analyzing the overall trajectory of this patient, not just what is happening during this hospital visit."

The [palliative care](#) staff still acts on clinician referrals in their daily rounds, but this model provides Harman with longer-range projections on people who might have been overlooked. On a typical day, she meets with the primary care team regarding two to three patients on the model's list. The cases that Harman selects are reported back to Shah's group so that they can monitor the algorithm's selection accuracy over time.

Harman has become attuned to the physical signs that a person is about to die. Breathing becomes irregular, with longer and longer pauses, the jaw dropping open with each breath. As the heart weakens, hands, feet and knees become mottled and cool to the touch. And there's the most profound moment, when there is only stillness. As a patient takes the last breath, Harman has her own ritual to usher them to the other side.

"I always say goodbye and thank you—sometimes out loud, sometimes not. I touch them, usually their hand or foot. I talk with the families. Sometimes they tell stories about the patient—usually funny ones. And I sit and listen for a spell. Even in the midst of such loss, families are welcoming. I express my sorrow for their loved one's death, though not always in words. I don't consider this the end; I believe that all of their souls carry on."

More information: Danton S. Char et al. Implementing Machine Learning in Health Care—Addressing Ethical Challenges, *New England*

Journal of Medicine (2018). [DOI: 10.1056/NEJMp1714229](https://doi.org/10.1056/NEJMp1714229)

Abraham Verghese et al. What This Computer Needs Is a Physician, *JAMA* (2017). [DOI: 10.1001/jama.2017.19198](https://doi.org/10.1001/jama.2017.19198)

Provided by Stanford University

Citation: Can machine learning bring more humanity to health care? (2018, September 20)
retrieved 3 May 2024 from

<https://medicalxpress.com/news/2018-09-machine-humanity-health.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.