Healthcare 8 (2020) 100465

Contents lists available at ScienceDirect

Healthcare

journal homepage: www.elsevier.com/locate/healthcare

Robots join the care team: Making healthcare decisions safer with machine learning and robotics

Adam L. Beckman [a,*], Julie A. Shah [b], Neel T. Shah [c,d], Sanchay Gupta [a]

[a] Harvard Medical School, Boston, MA, USA
[b] MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA
[c] Ariadne Labs at Brigham and Women's Hospital and the Harvard T.H. Chan School of Public Health, Boston, MA, USA
[d] Beth Israel Deaconess Medical Center, Department of Obstetrics and Gynecology, Boston, MA, USA

ARTICLE INFO

Keywords:
Health care
Health policy
Robotics
Artificial intelligence
Machine learning

In this interview, we spoke with Drs. Julie and Neel Shah. Julie Shah, PhD is an Associate Professor in the Department of Aeronautics and Astronautics at MIT and leads the Interactive Robotics Group of the Computer Science and Artificial Intelligence Laboratory. Before joining the faculty, she worked at Boeing Research and Technology on robotics applications for aerospace manufacturing. She has developed innovative methods for enabling fluid human-robot teamwork in time-critical, safety-critical domains, ranging from manufacturing to surgery to space exploration. Her group draws on expertise in artificial intelligence, human factors, and systems engineering to develop interactive robots that emulate the qualities of effective human team members to improve the efficiency of human-robot teamwork.

Neel Shah, MD, MPP is an Assistant Professor of Obstetrics, Gynecology and Reproductive Biology at Harvard Medical School, and Director of the Delivery Decisions Initiative at Harvard's Ariadne Labs. As an obstetrician-gynecologist at Beth Israel Deaconess Medical Center in Boston, Dr. Shah cares for patients at critical life moments that range from childbirth to primary care to surgery. As a scientist and social entrepreneur, he is a globally recognized expert in designing, testing, and spreading solutions that improve healthcare.

They live together in Cambridge, MA with their two toddlers and puppy. This interview was condensed and edited for clarity.

Adam Beckman (AB): Neel, you once Tweeted that you've "long thought that being the nurse in charge of a labor and delivery unit must be hardest job in health care." Can you walk us through why you said that and the subsequent collaboration Julie and you started?

Neel Shah (NS): In the Labor and Delivery Unit, you don't know when your patients are going to show up. You don't know how long each person is going to be laboring. And you don't know which one of them is going to get sick enough to suddenly need a significant resource, like the blood bank or an operating room.

So, the nurse in charge acts as the air traffic controller for the unit. This person figures out in real-time which staff nurse gets assigned to which patient, and which patient gets assigned to which bed. In addition, this person has to manage multiple inflows. You have people who are spontaneously in labor. You have a big antepartum unit with very sick people who at any moment can need urgent help. You have a postpartum unit where people can go, but if that backs up, people stay on the delivery unit. And you have NICU beds which can be depleted as well. Staffing for these units is really challenging, and safety is obviously critical.

Yet the state of the art at most hospitals across the country to solve this challenge–is a very experienced person with a pen and a piece of paper. From hospital-to-hospital, and from shift-to-shift within the same hospital, the person in that role approaches these incredible challenges in very different ways.

Julie Shah (JS): The job performed by that nurse is technically more difficult in terms of computational capacity than the job that an air traffic controller does. And they do it without any of the tools or decision support. It's incredible. One of my graduate students had calculated that

* Corresponding author. 25 Shattuck Street, Boston, MA, 02115, USA.
E-mail address: adam_beckman@hms.harvard.edu (A.L. Beckman).

https://doi.org/10.1016/j.hjdsi.2020.100465
Received 3 August 2020; Accepted 12 August 2020
Available online 6 September 2020

the mental tasks of the floor nurse would crash the world's fastest supercomputer on the basis of the number of futures that it would have to anticipate and create contingencies for.

Neel and I spend a lot of time thinking about when a computer is better at solving these kinds of things than a human being is, and also how to ensure humans are freed up to solve the problems that computers cannot. People are not well suited to solving scheduling problems or optimization problems in their head, particularly for problems that are well-structured. By contrast, the human brain is best focused on the unstructured, gray-zone kind of problems. When Neel and I started collaborating, the first task was trying to parse what types of tasks a computer could take on.

Right now the problems that these nurses are solving in their head are of a size that take a long time to solve with existing tools. This is part of the reason people are in these jobs through apprenticeship or years of practice and training. There's no textbook. They often do not know how their colleagues do it. There is no standard; there are no rules.

The challenge on our side was to develop new machine learning models that would be able to watch how people make these decisions in complex environments. Unlike many machine learning tools, you have to do the learning with relatively little data. We don't have an unending source of labelled images to learn from. And gathering data from human experts is very expensive. So the question was, "What can you do to address an ill-defined problem by learning from human experts and elaborating parts of the human decision-making strategy?"

Sanchay Gupta (SG): Increasingly in healthcare, people are incorporating 'machine learning' and 'artificial intelligence.' Julie, for those of us with a limited understanding of these technologies, what's something about them you think everyone in healthcare should understand? Neel, is there anything new you learned about these concepts from working with Julie and her team?

JS: The common perception of machine learning as a black box that will magically learn what you are doing and be able to potentially do it better is a bit unrefined. All aspects of the pipeline of improving these systems requires cultivating an understanding of the domain and the environment. It takes working with the experts to understand how we give the machine the right scaffolding to see our world. We need to help it understand how to piece together its observations of our world.

NS: To even get to the point of being able to write code to do this, one of Julie's graduate students basically lived like a resident. He hung out with our charge nurses and he shadowed me for the better part of two years.

JS: The point of that process was to understand what "features" these nurses are using in their decision making. It takes years for a machine learning expert to work with clinicians to do that translation for the machine. Ultimately, that's what it takes for the machine to do this job well.

We have also applied these machine learning models to emulating military experts performing missile defense tasks. For example, when multiple incoming missiles are heading to a ship, a naval officer's job is to figure out how to deploy their various decoys and counter mitigation resources. Like the resource nurse, there's no codified training procedure for how to do this. Some people in the simulation environment get really good at it. Other people are never good at it.

However, unlike the labor and delivery floor, in the military situation we have an objective measure to optimize for: survival of the ship. Because the military officers train in a simulation environment, it provided us an environment where we could fully encode the problem for a machine. The machine could quickly find solutions that were better than the best human experts.

That example just goes to show: There is potential for this type of human-machine collaboration to be better than people who are doing it alone. But we need to give the machine its starting point by showing it what our best people already know.

NS: The intuition of how this worked that helped me was that most human experts work pretty tacitly. If you're a world class athlete like Michael Jordan in his prime, you don't break down the steps. You just act intuitively. Michael Jordan can't tell you how to perform exactly like Michael Jordan. He can't walk off the court and be like, "This is how you do that." But right after he did something, he could tell you why he did it, which is the idea of how the technology works.

Julie's graduate student spent years embedded with us. Then he created a computer simulator, where the charge nurse could make assignments in different scenarios. Right after you did something in the simulator, the algorithm could try to back out the decision making process.

AB: We are curious to go through the specific phases of the project. What was the progression like?

JS: The first step was embedding the graduate student with Neel's team and the hospital. In parallel to that, we worked very closely with Neel and our collaborators at Beth Israel Deaconess Medical Center to develop a high fidelity simulation environment of the labor and delivery floor. We developed the machine learning models concurrently to collect data from nurses and doctors, playing a day in the life of their jobs in this simulation environment. And then that provided the data set that we use to train the learning model.

The final question for us was, "Could this be used on a real labor and delivery floor?" In the simulation environment, you control the full state of the world. You give the full state of the world to the machine, and it tells you its suggested next action. We decided to deploy a robot in order to answer the reality question.

We wanted to see if the robot could read the handwritten whiteboard on a real labor and delivery floor–and make a reasonable suggestion.

NS: The robot could predict the decision that you'd want to make. And it was pretty good. It would make suggestions, and the nurses and doctors agreed with the decision of the system 90% of the time. The idea was not to replace the experts, but to just deploy their bandwidth in an optimal way.

A nurse knows that if you do a lot of C-sections on Monday, the length of stay is always four days. So on Thursday the postpartum beds are going to be really backed up. But it takes bandwidth to have to keep track of that. That's the most basic example, but there are many similar ones.

If you take 90% of the little tasks away by adding a robot, you then have these nurses with two decades of deep expertise focused on the most safety critical things. Everyone is better off.

AB: What's another dysfunctional area of the healthcare system that you think is ripe to benefit from machine learning strategies and robots?

JS: I run an interactive robotics lab, and a growing area is service robots. I am talking about robots that roam your halls, deliver medications, deliver linens, help turnover rooms. And they're doing these support tasks that are meant to help move the flow of the hospital along and offload some of that work from people. These systems have been deployed to some extent, and then rolled back, repeatedly for 10 plus years. That's partially why you don't see them that often.

There's actually a fair amount of study in my field about why we don't see greater adoptance of these systems despite all of these potential benefits. One of the challenges is that these systems are generally not intelligent enough to understand the workflow of these complex environments. Someone has to task and schedule them. Which means in addition to those 15+ direct reports that resource nurse has, she now potentially has a fleet of robots that she's required to track and modify the schedules for.

Systems that are able to learn in the way people do–learn models of what's likely to happen in the future and make suggestions (i.e. when a room will need to be turned over or what needs to be delivered when)–make it much easier for a person to accept those suggestions than to explicitly task the systems. This class of techniques is an enabling technology for wider deployment of service robots and health care as well.

NS: In the abstract plane, I would say that any place where there's room to improve the reliability of what we do in health care, AI is likely to improve suffering in the near-term. There are a lot of things that experts think that they do, but because they're always acting tacitly and have to act in teams–where the information to act lives in multiple people's heads–we forget things.

The surgical safety checklist is a great example of this. Extremely basic stuff that we thought we were doing already–but the difference between doing it reliably and not may drop surgical mortality in half for every surgery on every continent.

On top of that, there are so many scheduling tasks in healthcare that could use help. I am going to make these numbers up, but if you have a 93% consensus at one of our tertiary medical centers, the hospital could be hemorrhaging money. But if you have a 98% census, the capacity may be unsafe. This basic task allocation thing matters–a lot.

SG: What do you see as some of the biggest obstacles to bringing these sorts of technologies and data analytics to improving healthcare in terms of quality or cost?

NS: We need more opportunities to have peoples' expertises come together. We need more people from the technology world embedded in our clinical environments and vice versa.

But there's another critical obstacle: We need to build trust in the system. Julie says that planes could fly themselves, but nobody would tolerate a person not being there. Julie showed me a couple of years ago a surgical assistant robot that does a small subset of the tasks that a surgical assistant does. But most of us are not ready to use that technology.

JS: In aerospace, we have a decades long study of human interaction with cockpit automation. It turns out to be very easy to manipulate someone's trust in a robot. For example, trust will vary depending on whether or not you give it a voice, and what specific voice you give it. You can change someone's propensity to accept advice from a system, and that's not necessarily coupled to the performance of the system. Even if you are fortunate to have a machine learning model that can express its confidence in its prediction, people don't know what to do with "80% confident." People still don't know whether to accept or reject the robot's advice.

NS: Another obstacle concerns translation. When thinking about how to report this project back to our clinical community, we thought: "It would be great if we could back out what the rules are that the robot has learned." It turns out that the robot doesn't speak English. We were left trying to parse these crazy decision trees in ones and zeros.

JS: And that's so funny, because from a robotics perspective, this project used technology that's more interpretable than other options, like a neural net. The robot was able to learn a decision tree for the nurse's decision making. But even that decision tree was not interpretable enough for a human being. Interpretability is a fundamental challenge in the machine learning community that's actively being worked on by a huge portion of the community.

AB: Julie, you've worked mainly in other industries like manufacturing and aeronautics. How is applying your work to healthcare different from those other industries?

JS: In many aspects, the research is similar. No matter the industry, you need to gain a detailed understanding of how the work is performed before you can enhance productivity. That said, a major difference is that doctors and nurses tend to have much less experience with robotics or standard automation than workers on a factory floor.

This created an openness in people's' minds of imagining the different ways that technology can enhance their work. In the media, the conversation is often centered around a fear of how systems will replace people. I'll note that in the near and medium term, these technologies are not going to replace large swaths of human work. Even in the most optimistic view, AI and machine learning technologies can do little pieces of many, many people's jobs but not replace whole jobs. Regardless, collaborators in manufacturing are sometimes afraid we will replace them.

By contrast, from our first interaction with the team on the obstetrics ward, this fear was not present. It was very clear that these nurses understand they are doing an extremely hard job. Their attitude was frankly, "Anything you can do to help would be great."

SG: Neel, you've done a great deal of advocacy through your organization Costs of Care. Can you tell us more about this organization? How do you balance your advocacy work with your work in innovation and technology?

NS: When I was in my third year of medical school, and rotating through the hospital for the first time, it felt like a veil got lifted. Many aspects of patient care were inspiring, others deeply disillusioning. I saw that as clinicians we make decisions all the time but have very little insight into how our decisions impact what the people in front of us have to pay for. I thought that clinicians ought to have a role in thinking about affordability given that we make the decision to end up on the bill. So, during my year doing a masters, I started an advocacy nonprofit called Costs of Care.

Ironically, about halfway into it, I joined the faculty to become a scientist. I had to begin to reconcile with the fact that scientists are supposed to be objective, whereas advocates are necessarily not. Along the way I've developed a much more liberal interpretation than some of my colleagues about where the lines are between the two. Impact requires science and advocacy to be coordinated. Science tells you that you're directionally correct, but it doesn't always help you get to where you want to be going.

In the last few years, I've tried to think strategically about both. Even on the scientific side, I've tried to pursue questions that I think are going to advance some strategic purpose. And I try to thoughtfully time what I have to say about those issues and build coalitions around them.

The lab I work with–Ariadne Labs–has a mission to get scalable solutions into the world. In healthcare, we tend to design a lot of products that don't have markets. But the goal is to sit down at some point and be able to say, "We spread that solution around in the world." To do that, you have to make sure the problem is visible. Let people see the scale and magnitude of the problem. Creating the market–the coalition that will spread the solution–often takes longer than creating the solution. In healthcare we often start with designing the solution and then spreading it, but you need to be doing both in parallel.

AB: Julie, on a more personal note, can you tell us a bit about your experience being a woman in a traditionally male-dominated field? What were some of the challenges you faced in STEM and what helped you overcome them?

JS: It's getting better, but there is a long way to go. My lab and I think a lot about how hard it is to bring people into a field unless you have visible role models. Whether at the undergraduate level, faculty, or industry, it's usually harder for women to find someone to look at and imagine what their path could be. It's human nature to want to do that. I've benefited greatly from a wonderful set of mentors, men and women, but the women were key. So, now I work hard to try and do that for others at all levels. There's a role for everybody to play, and it's getting better over time, which is encouraging.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research project we describe was funded by CRICO.