SRINIVASAN SURESH: Good morning. Thanks, Andrew. Thanks for the opportunity. This is a technology-based talk. So I'll be talking about health information technology firms and products, but I have no financial relationships with the industry and no conflict of interest. I'm on a council with [INAUDIBLE] Corporation in a purely advisory role.

A quick overview of what we'll discuss over the next 45 minutes or so-- I'd like to start with some definitions of key items in health informatics. Some use cases and successes. And these have come to light over the past one to two years. Literature review. Few examples of the applications of artificial intelligence and machine learning in adult medicine.

And I'm particularly proud to highlight two or four machine learning-based projects here at CHP. And it's good to know what's happening in other service industries outside of health care and how they use AI and ML. There certainly are some limitations. And we'll talk about it. And I have a few short videos-- about two minutes each-- to highlight some of the newer technologies.

I'd like to go over my credentials briefly. I'm not an expert in machine learning. Not an expert in AI. I'm not in authority neural networks. I do not have any publications in deep learning. Most folks in this field are trained at MIT or Stanford or Harvard-- not me.

I do not work for Google or Watson Health. I don't own a start-up. And most leaders in this field are either cardiologists or radiologists and intensivists and I'm not. So when you're wondering, how did this guy get up there?

I believe that is a role for clinicians, experienced physicians, and administrators-- leaders in this space. We need to part with the incredibly smart software engineers and data scientists by providing them with a clinical context to make these applications valuable. So that's what I see my role.

We'll do some key lingo. What is data? Data is a set of values of either qualitative or quantitative variables. Information, on the hand, is related to data and knowledge. It's basically the answer to a question of some kind.

Big data has two definitions. The theoretical definition is it's a collection of data sets that are large and complex that cannot be managed using standard database systems or traditional data application processes. A sort of working definition is when you have a data set which is more than 5 or 8 terabytes or bigger that comes under big data.

Informatics is basically the processing of information. It's capturing, storing, measuring, and analyzing information. Health informatics, to be specific, uses health information technology to improve health care via any combination of higher quality, higher efficiency, and new opportunities. Analytics-- again, a simple definition is data plus meaning-- is the discovery, interpretation, and communication of meaningful patterns in data. If you notice, the definition of health informatics has the words quality and efficiency in it.

Predictive analytics is a combination of a few fields-- statistical techniques from modeling, machine learning, and data mining to analyze current and historical data to make predictions about future or otherwise unknown events. Artificial intelligence is simply replication of human intelligence in computers. It deals with the theory and development of computer systems that have to perform tasks that we humans do sort of easily, like decision making and visual perception.

Machine learning is the ability of a machine to learn using large data sets instead of hard-coded facts. It gives computers the ability to learn without being explicitly programmed. Deep learning, on the other hand, is a type of machine learning where we train an artificial intelligence system to predict outputs given a set of inputs.

Deep learning software attempts to mimic our neocortex where our thinking occurs. The software learns, in a very real sense, to recognize patterns in digital representations of sounds and images the way we humans do. And finally, neural network is a computer system modeled on the human brain and the nervous system.

So let's talk about anatomy of a neural network. So you have a depiction there where you have an input layer. You can either have one or more hidden layers and then an output layer. So these are layers of neurons very similar to what we have in our brain. And they're all connected.

And let's take an example-- airplane ticket price estimation service. So the input for something like this would be-- this is a simplified version-- origin of the airport, destination of the airport, departure date, and airline. And even with these inputs, you can imagine that you'll have tons of data, depending on how many flights there are from the airport and is the departure date on a Saturday evening versus a Tuesday morning, et cetera.

The hidden layers perform the mathematical computations of our inputs. And finally, you get the output, which is the price. And once this is done, it has the ability to predict future prices.

One more step in the process-- the initial weights upset pretty randomly from the inputs. To do something like this, you need a pretty large data set. You need to know the details of every single flight between those two cities, the dates, and the prices so you can enter everything in and train the system. It's an iterative process to finally get to the real values. There is an item called a cost function, which is-- you can think of it as noise in this data-- which has to be minimized as you get to the final output.

And you can see some similarity between our neurons and the mathematical model. You have dendrites right on the left side. If you go from left to right, you have the cell nucleus and the axon, which carries impulses throughout our cell body, and finally, the output.

And the right side you could see a similar setup. You have axon from a neuron. And then these are the hidden layers. And then you have the output. So this is a biological neuron and its mathematical model.

An overview for AI and ML in pediatrics-- this is an excellent position paper by Charles Friedman. This is self-explanatory, but I want to talk a little bit about it. He talks about how information technology can augment human reasoning, as opposed to what the technology itself is capable of doing. So two corollaries from this simple theorem-- one, informatics is more about people than technology.

Number two, in order for this theorem to hold, the resource-- which you see with a computer icon-- must offer something which the brain doesn't have. And the brain here could be a person, could be an organization, a clinician, a researcher, a scientist-- could be anything. And the plus here is not to be read literally, in the sense of mathematical addition, but to convey the interaction between the person and the resource, the outcome of which is determined by what the information resource is capable of, as well as how the person elects to use it. Figure 2 is self-explanatory-- what informatics is not.

I'd like to highlight some studies which have been published recently using machine learning techniques. This one was interpreting gram stains using a neural network. Over 25,000 images from positive blood culture gram stains were collected. And from these, the system created 100,000 crops of four categories. These are not all the bloodstream infections, but the commoner ones.

The classification algorithm was then applied to 189 whole slides. And you can see the sensitivity and specificity were pretty high. The aim here was not to displace the technologist, but to augment the technologist's power so that he or she can read the images much faster if provided just the crops of those images instead of the whole slide.

Another example-- diagnostic assessment of deep learning algorithms for detection of lymph node metastasis in women with breast cancer. So the question is, what's the discriminative accuracy of deep learning algorithms compared to the diagnosis of pathologists? And they have found that seven deep learning algorithms showed greater discrimination than a panel of seven pathologists in this particular study.

And the takeaway from this is there is a possibility to expand this into the real clinical setting. Quite a few of the studies in machine learning literature are proof of concept or proof of principles. Some of these are not peer reviewed. So it's still early stages to take these and extrapolate it.

So based on those two slides, we now have a framework for machine learning in pathology. And thanks to Dr. Panotowitz for providing this slide. And to keep in the background-- that FDA last year approved the first whole slide imaging system for digital pathology. So when you talk about computational pathology, you have big data-- you have images together. They form the computer-aided diagnosis.

Then you have image analysis of lots and lots of images like the two previous examples. You have the convolutional neural networks plus machine learning leads to clinical decision support. This is just a broad framework. And every pathology AI or ML doesn't follow this, but it's good to have this.

Switching on fields, let's go to dermatology. So a convolutional neural network was trained using over 120,000 clinical images, which comprise of 2,000 skin diseases like acne, melanoma, moles, et cetera, just using pixels and disease labels as inputs. The performance was tested against 21 dermatologists on these two conditions. Keratinocyte carcinoma is the commonest skin cancer versus benign seborrhea. The next one is malignant melanoma, which is the most serious skin cancer, versus a benign nevus.

And the performance was on par with all tested experts-- slightly better, but pretty much in the same range. And this is kind of interesting because traditionally, when medical students and residents are taught about melanoma, there's a mnemonic, ABCD. A is asymmetric, B is borders that are uneven, C, the color can be patchy, and D, the diameter is greater than 6 millimeters. But apparently, not a lot of melanomas follow this ABCD rule. That prompted these researchers to look at it from a machine learning perspective and then achieve a pretty similar result to what experienced dermatologists would do.

Another landmark study, which was published a year ago, was from Google Research. And they had collaborated with some big names in UC Berkeley, UT Austin, and Brigham and Women's in figuring out how we can detect diabetic retinopathy. So they also had over 100,000 retinal images, which was created by 54 ophthalmologists. And 40 are ophthalmology residents. Had an algorithm which was validated and graded by US board-certified ophthalmologists.

The fundus exams were taken from the US, France, and India. So this deep learning algorithm had really high sensitivity and specificity for identifying diabetic retinopathy in retinal fundus photographs. So begs you the question, is that a roll where, in some parts of the world where you don't have easy access to ophthalmologists, can this technique be used to identify diabetic retinopathy?

Another interesting study which was presented at a conference-- I couldn't find the paper for this-- couple of months ago-- was AI-assisted endoscopy. And it caught the media's attention because the machine learning technique could predict the lesion in less than one second. So what they did was they magnified the view of a colorectal polyp by 500 times. Once you do that-- and these are high-end cameras, too-- you have 300 features which are analyzed against a trained set of 30,000 images. This was done on, again, 250 patients, who had about 300 polyps.

And compared to the final path report after resection-- and you could see the numbers-- pretty impressive-- sensitivity in the 90s, high positive and negative predictive values, and accuracy of 86%. So how does this help the provider or the patient? One, it's a real-time optical biopsy of polyps during colonoscopy. So this prevents unnecessary removal of non-neoplastic polyps. And two, allows complete resection of adenomatous polyps, thereby decreasing, hopefully, the onset of colorectal cancer and maybe even death from cancer.

Completely switching gears, I'd like to talk about a couple of projects that we have here at Children's. One of them is to figure out how we can apply machine learning techniques to predict 30-day all-cause readmissions. A shout out to my team. And this falls under the category of, how do we use quality implement techniques to improve transitions and reduce readmissions?

A quick background-- hospital readmissions pose a financial and reputation burden to a lot of hospitals. We have some CMS penalties which may not have hit us yet-- so far-- for readmissions. What if your hospitals publicly report this data? Huge difference, still, between adult and pediatrics. The readmission burden is higher in adults, and so are the penalties.

We felt there is a need to predict-- to develop an algorithm which can focus on inpatient pediatric care. So that was our goal-- to develop a predictive model which would provide a dynamic probability of a 30-day all-cause readmissions to providers in our hospital at point of care, where we have over 22,000 admissions now, annually. So how do we go about this? This was in 2014. So we use DRGs and ICD list to classify diagnosis.

The first step is to build a training data set, so we looked at our patients from a five-year period-- so that close to 100,000 discharges-- and I think 12,000 of those were avoidable readmissions. So what was different about those 12,000 versus the 88,000 who did not come back? So an algorithm was created. We just then validated against inpatient visits in 2012 and 2013 to refine it. And the way this was done is to use what's called a Bayesian network predictive model, which is actually not one of the newer ML techniques.

There are newer techniques to even make this better now. As you can see, we had a large number of cases and controls, both in the training data set as well as in the test data set. When we plotted to receive an operating characteristic curve, we had an AUC of 0.79. There are commercially available products with an AUC in the 0.6 and maybe low 0.7 range, but not in 0.79. And we have an option to look at specific diagnosis like seizures, asthma, and pneumonia.

We're calling the system a SHARP, which stands for System for Hospital Adaptive Readmission Prediction and management. And I'd like to share some results from silent mode, which was for a period of three months last year. And all of these are discharges which exclude planned readmissions. If you have somebody who needs chemo every three weeks, et cetera, those patients would be excluded from this. Let me walk through this table.

To go left to right, we had 3,000 discharges in those three months. Almost-- it was 50%/50% where the system categorized these patients at point of care discharges either low risk or high risk. And then if you follow the high risk patients, about 25% of them came back. The low risk-- only 7% came back. So even without doing any stats, you could see that there is some value to this.

And our overall readmission low is about 15%, so pretty close to it. What is interesting here is those 470 patients who actually came back within 30 days-- 369 of them were identified as high risk at the point of discharge. So that gives us an accuracy-- not a positive predictive value-- but an accuracy of 79%, which, again, is better than the commercially available products. We are proud to share this because this was totally homegrown. This was done in partnership with our Department of Biomedical Informatics and thanks to a generous donor and our Children's Hospital Foundation who helped us connect with the donor.

There is a potential to commercialize this. And similar children's hospitals like us could benefit from this, too. So now that you have an algorithm-- you have identified high risk patients-- then what? What do we do to actually implement it and potentially prevent readmission? So we did a pilot-- again, small numbers.

I'm not getting too excited with the results, here, but it's a start. So we did this pilot for two months in July and September-- we had-- July to September-- we had 256 discharges during that time. And all of these patients belong to a particular PCP group. So we had a control group and an intervention group-- not by design. This was a convenience sample.

We had 150 discharges in the control group where we did not do any intervention. And you can see 29% of them came back in the intervention group, where our intervention was phone calls by nurses four times during those 30-day periods. So pretty low-tech, but since we had a cohort list of high-risk patients, we were able to do that.

So a little bit of a busy slide, but I'll walk through it. And this is splitting the 106 patients-- whether we were able to call them once, twice, three, or four times. So these calls happened on day 2, day 7, day 10, day 14 post-discharge. And if you look at the last three columns, it's kind of interesting. Those 20 patients-- that 15 plus 5 who got at least three calls or more-- very little readmissions.

And we still have some work to do based on the feedback from the parents, who specifically said, I have a problem getting into this subspecialty clinic. Maybe we need to address that. And that would potentially prevent the two readmissions we had there. But it's exciting to know that out of the 60 discharges we have every day from our hospital, now we have a tool which can identify 16 or 18 or 20 high-risk discharges. So we can have our nurses call those 16 or 18 patients and target them with care coordination.

Peter Drucker is a famous management consultant, educator, and author. That's his quote, "If you can measure it, you can manage it." And somebody else said if you cannot automate the measurement of it, it's difficult to prove it. So we go with this. And what I showed now-- it is automated in our sonar EMR.

So we don't have to calculate it. It's not a separate website. It pulls in existing information in the EMR patient-related info and then provides a probability. These tools need to come up with a simple number, so this is a probability of the patient getting readmitted within 30 days.

Our second project, which is ongoing at an earlier stage, is application of machine learning techniques to detect catastrophic events in our CSCU. A slightly larger team-- and thanks to Dr. Munoz and Dr. Lopez, who are not here now, but they're still involved in the project. So what is this? This is a project called CWIN, Cardiac Warning Index, which has the potential to detect early-- some specific end events like emergent intubation, cardiac arrest, or CPR or ECMO in CSCU patients. The current warning systems we have are really not validated in ICU populations, and definitely not in patients with congenital heart disease.

So we have two phases for this-- a predictive analytics-- sort of an expert model, which we are done with-- and a machine learning component, which we are just starting. Our aim is to decrease morbidity, hospital length of stay, and cost of care. The mortality is already low, so I'm not sure whether this will actually improve mortality score or not. So in phase one, we looked at a basic model for single ventricle patients. And when you have patients with abnormal values-- I'm hoping you're able to read that.

So the three boxes in the bottom are systolic blood pressure, base excess, and oxygen saturation. And based on where those are, quickly the critical event-- probably a percentage-- comes up. And as you can see, this is for pictorial representation. If you have a good blood pressure, base excess, oxygen saturation, et cetera, the chances for a critical event is pretty high. Sorry, these are abnormal values.

And I'll show another one where the values are normal. And this is interesting because the oxygen saturation for these patients-- the ideal one would be 70% to 80%. So it's different than what we would have for patients without CHD, given their altered physiology post-op. I've shown three metrics here. And we are looking at two types of expert models.

One takes [INAUDIBLE] on 11 characteristics, including pulse ox, systolic, diastolic blood pressure, et cetera, and a bigger model, which takes a [INAUDIBLE] about 50 metrics. Phase two of this process is a machine learning component, which we're excited about just getting started, where we'll have a lot of information about the patient being analyzed by the types of machines we talked earlier, including telemetry and waveforms. And hopefully we'll find something there. There is an opportunity to expand this from just single ventricle patients to all CSU patients. And also, it's exciting that we could expand this to our other care sites-- maybe at least the expert model, if not the machine learning model.

Switching gears again-- area focus of patient access. And I'd like to thank Dr. Marino for providing these slides. Using GIS, what Dr. Marino did is she analyzed failed appointments at our primary care clinic in Oakland. So her questions were, do normal patients live in different zip codes than patients who complete their visits? And is there a correlation between no-show visits and poverty?

So she was able to come up with this. The red dot you see is the primary care center. And interestingly, it's not always the distance from the clinic which is an issue. You have some areas there which are reasonably close to the clinic but show a high no-show rate. And she was able to sort this out by zip codes.

Again-- the same concept-- the little black dot is the gap clinic. And then you have areas reasonably close with high no-show rates. This is with one clinic with about 10,000 to 20,000 patients. So imagine if we were able to do this-- what I showed now is not an example of machine learning. But if we were able to do this for every patient who comes to every one of our subspecialty clinics and tried to build a model that would border on machine learning, it has two advantages-- one, if we have a no-show risk score for a patient and the parent calls in, we may be able to intervene at that time.

We have to figure out how to tell the family, you have a high risk of not showing up, and then actually calling for an appointment. But the point is that's a high-risk patient. And if you have that information right then, you can do something with it. And number two, if you really have some high-risk patients who despite the nurse intervention are still not showing up-- I'm putting on my administrator hat now-- can we double book? Is it an option to actually use all of our capacity to see the optimal number of patients? So interesting area.

And this is one slide from Dennis Wallace, PhD. I attended his presentation a few months ago and it's fascinating. He's one of the leading autism researchers in the country. And what he has done is built a system where the parents take a video on their phones-- two to five minute videos-- of kids just playing around. And he has taken more than 10,000 of these videos.

But you need humans to watch these videos and then categorize into-- is a child joyful, sad, depressed, angry, and all these characteristics. You have to label those, which is the constraint. But based on that, he's able to get a 90%-plus sensitivity and accuracy in not only diagnosing autism, but also a few other behavioral developmental issues.

I'll switch over to speech recognition technology. And all of us deal with this. Now, there is a system-- this is not the future, it's present-- that automatically transcribes doctor-patient conversations. So I go into the ER, I see a patient, I spend five minutes in the room, I come out. And again, this is the proof of concept at this point.

But using recorders in the room that transcribes what I talk with the family and what the parent and the child told me and then converts into medicalese, so to speak. How did they do this? They looked at a [INAUDIBLE] of anonymized conversations representing 14,000 hours of speech to train the model. I read this paper and I wasn't able to figure out if they had any pediatric patients. They talk about caretakers and caregivers, but that could also be geriatric patients.

The key is you need a significant amount of data cleaning. There are multiple speakers at the same time, there's overlapping dialogue, the recorder was placed closer to the physician, not so much closer to the patient, there's a wide range of speech patterns, accents, et cetera, and a lot of casual conversations. You know, how was the Steelers game yesterday? You need to take off those kind of terms. It was interesting.

The model performed really well, and it had a 98% recall for drug names. That's not where it failed our performance, optimally. The errors actually occurred in casual conversations. The "mms" and the "ahs" is where the model couldn't do that well. But it's a start. And it's a pretty exciting area. I'll try to show a video of this.

[MUSIC PLAYING]

This was a different company and a product, what came up in the video, but it's a prototype. So there'll be more iterations and more proof of concept. But it's pretty exciting. And this, as all of us know, is a very practical solution to a lot of the problems we face. So this quote was made by Leonard Kish six years ago. He's a health IT strategy consultant.

What do you guys think the answer is? It's not a drug, I can tell you that. It's not a med. Patient engagement. And we all see this. We live this. And I'd like to show one video which addresses this.

SPEAKER 1: Health insurance is complicated. Here, take a look at your insurance card. What's your plan name? That one? Or that one? That one sounds familiar.

What about your member ID number? So many numbers. Is that it? What about your copay? It's not you. Everybody deals with this. And if we can't find basic information, how can we answer harder questions like, is this doctor in network?

Does my insurance even cover this? It's a headache, and not just for patients. Doctors struggle with bad information, too, and end up having to turn people away. It isn't healthy for anybody. That's why Zocdoc created Insurance Checker.

Just snap a picture of your card. And Insurance Checker decodes what's important so you can find a doctor that's actually in your network. Pick your doctor, and we'll make sure your insurance is valid, confirm all your details, and share the exact information the doctor needs without a single phone call. It's good for doctors.

It's good for patients. It's good for anybody who just wants their health care to make more sense. Download the Zocdoc app now to use Insurance Checker and book appointments with confidence.

SRINIVASAN SURESH: So a fair amount of artificial intelligence had to go into it. It's a cool video, but if you think about it, they had to build in a lot of numbers. As the person was indicating, different plans from the same payout or insurer, co-pays, physicians who may switch in and out of networks-- so all of this went in so that for the consumer, it's just a matter of scanning their card, and then to the point they could even book the appointment once they scanned it.

We talked about AI and ML. And sometimes they go together and sometimes they are distinct. So there's one quote here which I find useful by Dr. Lefkowitz. Machine learning is valuable when you have sufficient data represented. You'll typically need hundreds of thousands of X-rays or slides, if you're starting a machine learning project, of the kind of problem that needs to be solved.

The machine can then understand the data enough so it can make the same kinds of predictions and get the same results a person would, given the next set of inputs. Artificial intelligence, on the other hand, is about probability and risk. What is the likelihood of getting diagnosed with a certain disease? What is the probability of responding or failing to a particular therapy? And one interesting observation was when they had radiologists look at mammograms, they-- I'm generalizing a little bit-- they had rarely missed a false positive. If they said this was cancer, that was cancer.

If you look at machines, generally they are the other end. They're generally more conservative. And they don't do any false negatives. So there is an opportunity to use machines first so nothing gets by them, and then give a smaller workload, so to speak, to the radiologist, so that we were doing things right. We are giving the machine lots of images to look at-- what the machine is good at.

And we are giving a radiologist, whose accuracy may improve if he or she has to look at a smaller number-- because it's already pre-screened. And the false negatives are taken away. So that's another way to collaborate.

A few data points from the AI health care market-- there was a nice paper from Accenture last year. And they predict that key clinical health AI applications can potentially create $150 billion in annual savings in the next eight or nine years. That's a big number. So they broke it down.

The first one is robotic-assisted surgery and virtual nursing assistants, which is very interesting. I have a nice video to show about virtual nursing assistants and an opportunity to talk to the company, too, which does this. And then a few other areas-- connected medicine. There's one there which is clinical trial participant identifier.

And I'll just give one example-- Apple's research kit when it was introduced in Stanford a little over a year ago for a particular trial in cardiovascular medicine for adult patients. That research gets set up on a phone. They were able to enroll 10,000 patients in 26 hours. It took Stanford medical center the prior six months to enroll 10,000 patients. So you can see how it's not just the technology and the hardware of the phone, but the software and the machine learning techniques and the AI which can expedite clinical trials.

Smart machine triage is what the virtual nursing assistant is. And I'll share that video in a minute. It predicts a consumer's future health usage. And there was a nice survey where it went to health care execs, hospital presidents, and such, where 81% of them indicated that it's extremely or very important to offer their products or services through virtual assistants on messaging boards.

The survey also identified 72% of health organizations who indicated that they are already using this intelligent virtual assistance. I'm not sure if that number is accurate. It also bothered me. So are we part of the 28% minority that we are not using the virtual assistants?

SPEAKER 2: Our app can be a part of their lives every day. So our avatar, Molly-- she can check in with the patient every single morning-- more than once a day, if that's needed. She can ask the patients how they're doing.

She can collect really important information, such as their weight or their blood pressure, glucose levels-- ask them relevant questions to whatever condition they might have. And all this information can be funneled back to the clinician seamlessly. So the clinician can just be alerted to which patients are actually, possibly having an issue that day.

MOLLY: Your weight is 190.9 pounds. You did not gain any weight. That's good. Now let's take your blood pressure. Please make sure the cuff is on by pressing the orange button on this side--

SPEAKER 2: Because we have this avatar that talks to you and you can talk back to, it makes people way more engaged because patients have told me it's comforting to have her. It's like someone's holding my hand. And they would rather talk to her than just use an app on their phone. Because of that engagement level, they're more likely to do these types of things to keep themselves healthier.

MOLLY: Blood pressure is 149 over 94. Your pulse is 81 beats per minute. Your blood pressure is a bit high today. We will keep monitoring that.

Now that we are done taking your vitals, I want to ask you more about how you are feeling. If you are experiencing shortness of breath, is it worse than normal? Anything else you want to tell me?

SPEAKER 3: No.

MOLLY: Because of your results this morning, you need to check in with the nurse to see if there are any adjustments you need to make to your clinical routine. It's outside of regular office hours, but I can still connect you with someone. Would you like me to do so?

All right. A nurse will give you a call back on the next business day to see how you are doing. Thanks for your time and I will talk to you tomorrow. Goodbye.

SRINIVASAN SURESH: And while they surveyed a lot of these users-- and they were all adult patients-- they actually preferred talking to this virtual nursing assistant. Talking instead of doing this-- we are just an app on the phone. It's really interesting. And I've seen the demo of this. And it's very close to talking to a real nurse. And they're getting better at this.

AI and the diagnosis of pneumonia. And this came from IEEE, which is a pretty reputed organization. It stands for Institute of Electrical and Electronics Engineers. The reason I'm putting this up is the moment-- this is some shock value. When you see this and it says that Stanford algorithm can diagnose pneumonia better than radiologists, you feel like, is that really true? Let me read the paper.

And they also had this little icon, AI versus doctors, which is, again, for shock value. So the paper actually-- the actual title is Radiologist Level Pneumonia Detection. They did not say that the machine beat the radiologist. But it's really interesting. So what these guys did was they had a system called CheXNet.

And that is a publicly available database of X-rays called ChestX-ray14. These are all adult patients, where they had over 100,000-- only frontal views-- AP or PA views of chest X-ray. And from 30,000 patients, 14 thoracic diseases-- pneumonia, pneumothorax, fibrosis, cardiomegalie was counted as a disease. So they had 14 diseases. And the key here was to say whether it was pneumonia or not pneumonia.

And CheXNet achieved-- it's called an F1 score. So it's not just a positive predictive value or accuracy. But the score was higher than for experienced radiologists on a test that are 420 X-rays. There are some limitations. Only frontal X-rays were available. But they had a study to quote-- that only 15% of pneumonias are picked up by doing the additional lateral chest X-ray.

And two, this was just a machine learning study. So neither the system nor the radiologist had clinical information. And even going back to the original title, I wouldn't say this was diagnosis of pneumonia. This was either positive or negative identification of pneumonia on an X-ray. And as clinicians, obviously, we would take history and other factors into mind while making a diagnosis of clinical pneumonia.

The interesting part is, if you remember the picture from a few minutes ago where you had the input layer, the hidden layers, and the output layer, we had one or two for depiction. This one, they had 121 hidden layers, which means this is some serious computational power, to look at I don't know how many features on these X-rays so that you go through every single level and finally get to the output. So every time you move through a lot of hidden layers, your accuracy is heightened.

And even looking at the number of X-rays they had and the success rate, it's a pretty fascinating study. Another example where the machine learning software assists with diagnosis on an iPhone-- the quantity is not very clear, but I'll try to read from the slide. And this one says try to go really close to the rash and take a picture to get a good focus. And then it's analyzing the rash.

The interesting part is that it's a new technology called Core ML, C-O-R-E M-L, where this picture does not go to the cloud. The actual machine learning software lives on our phone. So it analyzes this. The database is 32,000 high-definition skin images-- various rashes. And gives you a-- so just a diagnosis.

This is a clinician-only app. This is not consumer facing. And the clinician can put in a few history or physical exam points to refine it and even get some differential diagnosis. So this one says blanching rash. So it's probably nothing pathologic. But then you have close differential diagnosis like Lyme disease, [INAUDIBLE], et cetera.

They got a lot of mileage because Tim Cook, the CEO of Apple, mentioned this company in their quarterly earnings call, which is a big deal. And it still needs a lot of further study as to how this can work well. But the company went one step ahead. Now the technology is trickling down to consumers.

They have a patient-facing app in development. And it's primarily for education and triage. This is not for the patient or the parent to self-diagnose a dermatologic condition, but at least put them into buckets as to, should I go to the ER for this or can I call my pediatrician on Monday or get a dermatologist appointment.

AI and social media. Facebook created an AI tool that can prevent suicide but won't talk about how it works. This is the actual title. It's really fascinating that they have the talent to look at-- and obviously the data-- to look at numerous Facebook postings. And it's not just the keywords, it was when they were posted-- was it related after this patient saw a movie or not.

So they can deduce a lot of important information from it. And they dispatch first responders for wellness check. So this is a little iffy. I'll give you that. I've been reading about this to find out if you can get more information.

But academics and researchers may not jump onto this if they don't know what the algorithm is and what is the false positive and the false negative rate. Here's a lot of examples. So what are the policymakers doing? So they passed a bill. Senator Cantwell from my state of Washington-- she has pushed a bill, which you can look online, to set up a special advisory committee, which will be multi-disciplinary, to study and predict the impact AI will have on society.

It's called A FUTURE of AI. And the FUTURE actually stands for something-- Fundamentally Understanding Usability and Realistic Evolution. It's been deferred to the Commerce Committee for next steps. And to quote Senator Cantwell, "We expect that AI will be an incredibly transformative force for growth and productivity. We need to be ready for it."

AI and first aid. This was really interesting-- just happened a few weeks ago. Mayo Clinic introduced this first aid skill for Amazon Alexa. So most of you may have used Alexa or Google Home. It's a hands-free way to access first aid information. It's voice-driven self-care.

So you have to say, Alexa, open Mayo First Aid. It's for minor conditions. What do I do when I get a simple cut, simple burn? But the interesting part is it is Mayo content. It is whatever content Mayo uses on their website, on their mobile platform, on their portal. And the same content is now delivered through a new medium.

And some of you may have read this. Siddhartha Mukherjee is an oncologist and an amazing writer. So both of these are the same articles. And sometimes they have two different titles in *The New Yorker.* It was "AI Versus MD" and "The Algorithm Will See You Now."

So he talks about how deep learning systems have outperformed human experts. The word diagnosis is a Greek word from knowing apart. So he feels that machine learning algorithms similarly become better at such knowing apart, partitioning and distinguishing moles from melanomas. He talks a little bit about what are the legal liabilities if the machine makes a wrong prediction. And hopefully the policymakers will figure that out.

And he goes into detail about comparing how a child learns versus how the machine learns. I'll give an example. When a child learns that this is a dog or this is not a dog, it's just by intuition. The parents tell them, no, that's a wolf, it's not a dog. That's a rat, it's not a dog. And by making some errors, the child learns.

And literally, the neural networks-- that's how they are trained. They are also trained, now, to figure out dog versus muffin. This is hard. It really is hard. Again, as humans, we would know. I hope.

But the machines-- this is like the acid test for machine learning algorithms. Can they differentiate dogs versus muffins? And there a few other things, which you can Google and find out. But it's fascinating. This literally came out five days ago.

Google is using 46 billion data points to predict the medical outcomes of hospital patients. This is not peer reviewed. It came out as an article, but it's not peer reviewed yet. Claims of better accuracy than existing software-- whether the patient will die in a hospital, be discharged, and the final diagnosis.

200,000 adults-- more than 46 billion data points from two reputed medical centers, UCSF and Chicago. So early stages-- somebody from Google Cloud whom I met forwarded this to me and said, so what about pediatrics? So it would be nice if we could do something like this.

Switching gears a little bit, I want to convey that to do a lot of these projects and to sustain it and to publish it, you need significant organizational commitment. We are fortunate here to have both hospital as well as department leadership when we do projects like a readmission risk prediction, as well as the cardiac warning index, to positively impact patient care. And for those of you who have not seen our 10-year strategic plan, one of the key goals is to reinvent primary care with a focus on population health and chronic disease.

And I believe that we will be using some machine learning techniques. We're actually working with a start-up on the west coast who is looking at Dr. Stacy Cook's patients and some of her patients who are high health care utilizers. And feels confident that he can predict future high health care utilizers in our area.

The value of predictive analytics-- so this was by a former Surgeon General who said we could save full lives. He's not talking about reducing morbidity. He's talking about saving lives with better data-driven care coordination and follow-up after a hospital stay for a psychiatric episode. And we may not have these numbers in pediatrics, but if we use all these techniques for [INAUDIBLE] better patient experience, it'll be worth the investment.

For the trainees in the audience, I want to put this up. When we have these opportunities and we try to figure out, should we invest or not, the value equation comes into play. It's quality versus cost or as some people like to say, outcomes divided by cost. And you can break down quality into outcomes plus safety plus service, and even more-- one level down-- service and dissatisfaction plus access. It's a theoretical equation, but I find it useful to evaluate projects as a whole.

A few more slides I'll mention. One is personalized medicine, where University of Pittsburgh is a key part of this project, which was started by President Obama, where the aim is to collect clinical administrative and genomic data to provide personalized care. The aim is to enroll 1 million human beings in the US in the next five to six years. And Pitts' ask is 155,000 lives. We will be getting into this project, Children's, by the end of this year or early 2019.

And the aim is to enroll about 20,000 to 40,000 children to represent the population in this area. And this will be a machine learning-based project. I'd like to show this slide to indicate why being mobile-centric in AI and ML certainly brings down the cost per unit of service.

We are still very patient-centric and, to a certain extent, ambulatory-centric, but we need to move towards the right to be more mobile-centric. Another slide from Gartner, which gives you a nice picture of how analytics has evolved. And we do a lot of work on the left side of the slide-- descriptive and diagnostic. But once you get into ML and AI projects, then you go more into the insight and the foresight.

One word about the Pittsburgh study, which Dr. [INAUDIBLE] is championing. It's called Framingham for Kids, for lack of a better explanation. And the plan is to investigate correlates of health in Allegheny County birth cohorts. We'll be collecting a lot of data. We'll have to address that acquisition storage analysis and governance issues.

Again, will be a big data project. And we're planning to start a year from now. So how do we do this? We recruit talent. Henry is here. He's a PhD with knowledge in machine learning and biomedical data science. Those are the languages he knows. I don't.

Now the other side of the coin-- a nice viewpoint in *JAMA* on the unintended consequences of machine learning in medicine. One is this term deskilling. Does that reduce the physician's skills as measured by traditional measures? It probably does. And they had two small studies.

One showed a decrease in diagnostic sensitivity when readers were also presented with the information what the computer aided diagnosis was. Another one was the study of internal medicine residents, which showed that the resident showed a decrease in diagnostic accuracy when EKGs were annotated with inaccurate computer-aided diagnosis. I don't know why you would do that, but they tried it and then figured out the residents were getting it wrong.

Quickly, I'll show a few examples of AI in other service industries. The automobile industry-- previously, insurance companies had to rely on our gender, where we live, our previous crash history, et cetera. But now you have apps which are on the phone. You don't need any adapter. We leave the phone in the car and you're driving.

It uses GPS and the phone's accelerometer to look at are you a good driver or not. And it brings up dashboards. It's a pretty scientific tool, actually, because they look at five factors-- phone use while driving, speeding, accelerating, cornering, and braking. And based on this information, they were able to show there was a 31% improvement after using the app.

I find this useful because this is like two-year-old information. It probably has gotten better. But the data comes from 25,000 drivers with over 30 million miles tracked. So it's a pretty big corpus of information they have.

Completely switching gears to the entertainment industry in Hollywood, where this advanced analytics company called Legendary-- what they're trying to do is-- traditionally, moviegoers are classified into four quadrants-- adult males, not adult males, adult females, not adult females. That's how Hollywood was targeting their marketing campaign, et cetera-- just four quadrants. If you take the whole country's population, 80 million each. Their goal is to get to a microsegment the other way-- 80 million groups of four. They are down to groups of 600 now-- 500,000 groups of 600 people who they have figured out that, we need to send information on Godzilla to this group and not to this group, based on tremendous amounts of publicly available data.

Let's skip this and go to conclusions so we'll have some time for questions. AI can create value in four ways. One, it can predict disease, identify high-risk patient groups like we talked about, and launch prevention therapies. Can automate and optimize hospital operations, automate diagnostic tests, and make them faster and more accurate. It can predict cost more accurately and focus on patients' risk reduction.

And finally, provide some therapies and drug formulations under the umbrella of precision medicine. To do that, we need to be very data driven. I like this quote, so I'll mention it.

And professor Deming was a very famous management consultant, engineer, and statistician. We have an opportunity, locally. The Pittsburgh Health Data Alliance has funded grants based on advanced analytic techniques and how to apply data to improve patient care.

So my last slide-- impact of AI in health care. I strongly believe they are here to stay. This is a very transformative technology. It's in its infancy. It's also pretty disruptive. So we should be careful as to how we deal with it.

I like this line which was in the Google article where they looked at the retinal fundus pictures. It was not even in quotes, it was just a line in the article. But it caught my attention. "A machine will make the same prediction on a specific image every time." That is consistency of interpretation. I don't think we humans can do it.

And I'm not even talking about integrator variability. If you give me a picture of a sheet of paper with 10 rashes and I diagnose them with whatever I think it is and you give the same paper to me a week from now, a month from now, will I do it right? And will I get all my rights right and my wrongs wrong? I'm not sure. So that's where I think machines can augment our clinical reasoning.

Collaboration amongst clinicians and between clinicians and technologists are very important. There's a quote from Bill Gates. "We always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next 10." I still feel that the best medical technology of all time is the stethoscope because it takes me closer to the patient. I've been lucky to practice medicine without having to document on a computer while I'm in the patient's room. And I hope to continue that way.

That absolutely does not mean I will not harness the power of all this technology, but for the trainees in the audience, I think there is a balance to strike-- and it's not hard-- where we can use all this technology and still practice old-time medicine, for lack of a better word. And a final quote-- it's not about man versus machine or woman versus machine, but the quote I heard was, "Physicians who embrace or use AI may replace physicians who do not." Thank you.