

[CLICK]

**DAVID SONTAG:** So welcome to spring 2019 Machine Learning for Healthcare. My name is David Sontag. I'm a professor in computer science. Also I'm in the Institute for Medical Engineering and Science. My co-instructor today will be Pete Szolovits, who I'll introduce more towards the end of today's lecture, along with the rest of the course staff.

So the problem. The problem is that healthcare in the United States costs too much. Currently, we're spending \$3 trillion a year, and we're not even necessarily doing a very good job. Patients who have chronic disease often find that these chronic diseases are diagnosed late. They're often not managed well. And that happens even in a country with some of the world's best clinicians.

Moreover, medical errors are happening all of the time, errors that if caught, would have prevented needless deaths, needless worsening of disease, and more. And healthcare impacts all of us. So I imagine that almost everyone here in this room have had a family member, a loved one, a dear friend, or even themselves suffer from a health condition which impacts your quality of life, which has affected your work, your studies, and possibly has led to a needless death.

And so the question that we're asking in this course today is how can we use machine learning, artificial intelligence, as one piece of a bigger puzzle to try to transform healthcare. So all of us have some personal stories. I myself have personal stories that have led me to be interested in this area. My grandfather, who had Alzheimer's disease, was diagnosed quite late in his Alzheimer's disease.

There aren't good treatments today for Alzheimer's, and so it's not that I would have expected the outcome to be different. But had he been diagnosed earlier, our family would have recognized that many of the erratic things that he was doing towards the later years of his life were due to this disease and not due to some other reason.

My mother, who had multiple myeloma, a blood cancer, who was diagnosed five years ago now, never started treatment for her cancer before she died one year ago. Now, why did she die? Well, it was believed that her cancer was still in its very early stages. Her blood markers

that were used to track the progress of the cancer put her in a low risk category. She didn't yet have visible complications of the disease that would, according to today's standard guidelines, require treatment to be initiated.

And as a result, the belief was the best strategy was to wait and see. But unbeknownst to her and to my family, her blood cancer, which was caused by light chains which were accumulating, ended up leading to organ damage. In this case, the light chains were accumulating in her heart, and she died of heart failure.

Had we recognized that her disease was further along, she might have initiated treatment. And there are now over 20 treatments for multiple myeloma which are believed to have life-lengthening effect. And I can give you four or five other stories from my own personal family and my friends, where similar things have happened. And I have no doubt that all of you have as well.

So what can we do about it is the question that we want to try to understand in today's course. And don't get me wrong. Machine learning, artificial intelligence, will only be one piece of the puzzle. There's so many other systematic changes that we're going to have to make into our healthcare system. But let's try to understand what those AI elements might be.

So let's start in today's lecture by giving a bit of a background on artificial intelligence and machine learning in healthcare. And I'll tell you why I think the time is right now, in 2019, to really start to make a big dent at this problem. And then I'll tell you about-- I'll give you a few examples of how machine learning is likely to transform healthcare over the next decade. And of course we're just guessing, but this is really guided by the latest and greatest in research, a lot of it happening here at MIT.

And then we'll close today's lecture with an overview of what's different, what's unique about machine learning healthcare. All of you have taken some machine learning course in the past, and so you know the basics of supervised prediction. Many of you have studied things like clustering. And you're certainly paying attention to the news, where you see news every single day about Google, Facebook, Microsoft's latest advances in speech recognition, computer vision, and so on.

So what's really different about trying to apply these techniques in the healthcare domain? The answer is that there's a huge amount of difference, and there are a lot of subtleties to doing machine learning right here. And we'll talk about that throughout the whole entire semester.

So to begin, this isn't a new field. Artificial intelligence in medicine goes back to the 1970s or sometime even in the '60s. One of the earliest examples of trying to use artificial intelligence for diagnosis was this MYCIN system developed at Stanford, where the goal was try to identify bacteria that might cause infection and then to try to guide what would be the appropriate therapy for that bacteria.

Now, it was found that this algorithm, this machine learning, this simple AI algorithm was able to propose a good therapy in 69% of cases, which at the time was better than the best or very good infectious disease experts. Now, it also had some other elements of it which were quite interesting. So those of you who've been following the latest in chat bots might find this a little prescient.

So there's a dialog interface where the computer says, I am ready. The computer clinician responds, this is a 26-year-old male patient. The computer tests its understanding of what's going on. It says, the age of the patient is 26. The sex of the patient is male. The clinician writes, five days ago, he had respiratory tract infections. The computer asks what is his name. Joe.

The computer responds, my understanding is the name of the patient is Joe. Respiratory tract is one of the symptoms the patient had. Then the clinician writes, a couple of days before the admission, he had malaise, which is general tiredness. And the computer responds, please give me a date of admission. The clinician responds, March 12, 1979, and the computer again confirms that it's understood appropriately. And this is the preface to the later diagnostic stages.

So the ideas of how AI can really impact medicine have been around a long time. Yet these algorithms which have been shown to be very effective, even going back to the 1970s, didn't translate into clinical care. A second example, oh so equally impressive in its nature, was work from the 1980s in Pittsburgh, developing what is known as the INTERNIST-1 or Quick Medical Reference system.

This was now used not for infectious diseases, but for primary care. Here one might ask, how can we try to do diagnosis at a much larger scale, where patients might come in with one of hundreds of different diseases and could report thousands of different symptoms, each one giving you some view, noisy view, into what may be going on with a patient's health.

And at a high level, they modeled this as something like a Bayesian network. It wasn't strictly a Bayesian network. It was a bit more heuristic at the time. It was later developed to be so. But at a high level, there were a number of latent variables or hidden variables corresponding to different diseases the patient might have, like flu or pneumonia or diabetes.

And then there were a number of variables on the very bottom, which were symptoms, which are all binary, so the diseases are either on or off. And here the symptoms are either present or not. And these symptoms can include things like fatigue or cough. They could also be things that result from laboratory test results, like a high value of hemoglobin A1C.

And this algorithm would then take this model, take the symptoms that were reported for the patient, and try to do reasoning over what action might be going on with that patient, to figure out what the differential diagnosis is. There are over 40,000 edges connecting diseases to symptoms that those diseases were believed to have caused. And this knowledge base, which was probabilistic in nature, because it captured the idea that some symptoms would only occur with some probability for a disease, took over 15 person years to elicit from a large medical team.

And so it was a lot of effort. And even in going forward to today's time, there have been few similar efforts at a scale as impressive as this one. But again, what happened? These algorithms are not being used anywhere today in our clinical workflows.

And the challenges that have prevented them from being used today are numerous. But I used a word in my explanation which should really hint at it. I used the word clinical workflow. And this, I think, is one of the biggest challenges. Which is that the algorithms were designed to solve narrow problems. They weren't necessarily even the most important problems, because clinicians generally do a very good job at diagnosis. And there was a big gap between the input that they expected and the current clinical workflow.

So imagine that you have now a mainframe computer. I mean, this was the '80s. And you have a clinician who has to talk to the patient and get some information. Go back to the computer. Type in a structured data, the symptoms that the patient's reporting. Get information back from the computer and iterate. As you can imagine, that takes a lot of time, and time is money. And unfortunately, it prevents it from being used.

Moreover, despite the fact that it took a lot of effort to use it when outside of existing clinical workflows, these systems were also really difficult to maintain. So I talked about how this was

elicited from 15 person years of work. There was no machine learning here. It was called artificial intelligence because one tries to reason in an artificial way, like humans might. But there was no learning from data in this.

And so what that means is if you then go to a new place, let's say this was developed in Pittsburgh, and now you go to Los Angeles or to Beijing or to London, and you want to apply the same algorithms, you suddenly have to re-derive parts of this model from scratch. For example, the prior probability of the diseases are going to be very different, depending on where you are in the world.

Now, you might want to go to a different domain outside of primary care. And again, one has to spend a huge amount of effort to derive such models. As new medicine discoveries are made, one has to, again, update these models. And this has been a huge blocker to deployment.

I'll move forward to one more example now, also from the 1980s. And this is now for a different type of question. Not one of how do you do diagnosis, but how do you actually do discovery. So this is an example from Stanford. And it was a really interesting case where one took a data-driven approach to try to make medical discoveries.

There was a database of what's called a disease registry from patients with rheumatoid arthritis, which is a chronic disease. It's an autoimmune condition, where for each patient, over a series of different visits, one would record, for example, here it shows this is visit number one. The date was January 17, 1979. The knee pain, patient's knee pain, was reported as severe. Their fatigue was moderate. Temperatures was 38.5 Celsius.

The diagnosis for this patient was actually a different autoimmune condition called systemic lupus. We have some laboratory test values for their creatinine and blood nitrogen, and we know something about their medication. In this case, they were on prednisone, a steroid.

And one has this data at every point in time. This almost certainly was recorded on paper and then later, these were collected into a computer format. But then it provides the possibility to ask questions and make new discoveries.

So for example, in this work, there was a discovery module which would make causal hypotheses about what aspects might cause other aspects. It would then do some basic statistics to check about the statistical validity of those causal hypotheses. It would then

present those to a domain expert to try to check off does this make sense or not.

For those that are accepted, it then uses that knowledge that was just learned to iterate, to try to make new discoveries. And one of the main findings from this paper was that prednisone elevates cholesterol. That was published in the *Annals of Internal Medicine* in 1986.

So these are all very early examples of data-driven approaches to improve both medicine and healthcare. Now flip forward to the 1990s. Neural networks started to become popular. Not quite the neural networks that we're familiar with in today's day and age, but nonetheless, they shared very much of the same elements.

So just in 1990, there were 88 published studies using neural networks for various different medical problems. One of the things that really differentiated those approaches to what we see in today's landscape is that the number of features were very small. So usually features which were similar to what I showed you in the previous slide. So structured data that was manually curated for the purpose of using in machine learning.

And there was nothing automatic about this. So one would have to have assistants gather the data. And because of that, typically, there were very small number of samples for each study that were used in machine learning.

Now, these models, although very effective, and I'll show you some examples in the next slide, also suffered from the same challenges I mentioned earlier. They didn't fit well into clinical workflows. It was hard to get enough training data because of the manual efforts involved.

And what the community found, even in the early 1990s, is that these algorithms did not generalize well. If you went through this huge effort of collecting training data, learning your model, and validating your model at one institution, and you then take it to a different one, it just works much worse. OK? And that really prevented translation of these technologies into clinical practice.

So what were these different domains that were studied? Well, here are a few examples. It's a bit small, so I'll read it out to you. It was studied in breast cancer, myocardial infarction, which is heart attack, lower back pain, used to predict psychiatric length of stay for inpatient, skin tumors, head injuries, prediction of dementia, understanding progression of diabetes, and a variety of other problems, which again are of the nature that we see about, we read about in the news today in modern attempts to apply machine learning in healthcare.

The number of training examples, as mentioned, were very few, ranging from 39 to, in some cases, 3,000. Those are individuals, humans. And the networks, the neural networks, they weren't completely shallow, but they weren't very deep either. So these were the architectures they might be 60 neurons, then 7, and then 6, for example, in terms of each of the layers of the neural network.

By the way, that sort of makes sense given the type of data that was fed into it. So none of this is new, in terms of the goals. So what's changed? Why do I think that despite the fact that we've had what could arguably be called a failure for the last 30 or 40 years, that we might actually have some chance of succeeding now.

And the big differentiator, what I'll call now the opportunity, is data. So whereas in the past, much of the work in artificial intelligence in medicine was not data driven. It was based on trying to elicit as much domain knowledge as one can from clinical domain experts. In some cases, gathering a little bit of data. Today, we have an amazing opportunity because of the prevalence of electronic medical records, both in the United States and elsewhere.

Now, here the United States, for example, the story wasn't that way, even back in 2008, when the adoption of electronic medical records was under 10% across the US. But then there wasn't an economic disaster in the US. And as part of the economic stimulus package, which President Obama initiated, there was something like \$30 billion allocated to hospitals purchasing electronic medical records.

And this is already a first example that we see of policy being really influential to create the-- to open the stage to the types of work that we're going to be able to do in this course today. So money was then made available as incentives for hospitals to purchase electronic medical records. And as a result, the adoption increased dramatically. This is a really old number from 2015 of 84% of hospitals, and now today, it's actually much larger.

So data is being collected in an electronic form, and that presents an opportunity to try to do research on it. It presents an opportunity to do machine learning on it, and it presents an opportunity to start to deploy machine learning algorithms, where rather than having to manually input data for a patient, we can just draw it automatically from data that's already available in electronic form.

And so there are a number of data sets that have been made available for research and

development in this space. Here at MIT, there has been a major effort pioneered by Professor Roger Mark, in the ECS and Institute for Medical Engineering department, to create what's known as the PhysioNet or Mimic databases.

Mimic contains data from over 40,000 patients and intensive care units. And it's very rich data. It contains basically everything that's being collected in the intensive care unit. Everything from notes that are written by both nurses and by attendings, to vital signs that are being collected by monitors that are attached to patients, collecting their blood pressure, oxygen saturation, heart rate, and so on, to imaging data, to blood test results as they're made available, and outcomes. And of course also medications that are being prescribed as it goes.

And so this is a wealth of data that now one could use to try to study, at least study in a very narrow setting of an intensive care unit, how machine learning could be used in that location. And I don't want to under-emphasize the importance of this database, both through this course and through the broader field. This is really the only publicly available electronic medical record data set of any reasonable size in the whole world, and it was created here at MIT. And we'll be using it extensively in our homework assignments as a result.

There are other data sets that aren't publicly available, but which have been gathered by industry. And one prime example is the Truven Market Scan database, which was created by a company called Truven, which was later acquired by IBM, as I'll tell you about more in a few minutes.

Now, this data-- and there are many competing companies that have similar data sets-- is created not from electronic medical records, but rather from-- typically, it's created from insurance claims. So every time you go to see a doctor, there's usually some record of that that is associated to the billing of that visit. So your provider will send a bill to your health insurance saying basically what happened, so what procedures were performed, providing diagnoses that are used to justify the cost of those procedures and tests.

And from that data, you now get a holistic view, a longitudinal view, of what's happened to that patient's health. And then there is a lot of money that passes behind the scenes between insurers and hospitals to corporate companies, such as Truven, which collect that data and then resell it for research purposes. And one of the biggest purchasers of data like this is the pharmaceutical industry.

So this data, unfortunately, is not usually publicly available, and that's actually a big problem,



both in the US and elsewhere. It's a big obstacle to research in this field, that only people who have millions of dollars to pay for it really get access to it, and it's something that I'm going to return to throughout the semester. It's something where I think policy can make a big difference.

But luckily, here at MIT, the story's going to be a bit different. So thanks to the MIT IBM Watson AI Lab, MIT has a close relationship with IBM. And fingers crossed, it looks like we'll get access to this database for our homework and projects for this semester.

Now, there are a lot of other initiatives that are creating large data sets. A really important example here in the US is President Obama's Precision Medicine Initiative, which has since been renamed to the All of Us Initiative. And this initiative is creating a data set of one million patients, drawn in a representative manner, from across the United States, to capture patients both poor and rich, patients who are healthy and have chronic disease, with the goal of trying to create a research database where all of us and other people, both inside and outside the US, could do research to make medical discoveries.

And this will include data such as data from a baseline health exam, where the typical vitals are taken, blood is drawn. It'll combine data of the previous two types I've mentioned, including both data from electronic medical records and health insurance claims. And a lot of this work is also happening here in Boston.

So right across the street at the Broad Institute, there is a team which is creating all of the software infrastructure to accommodate this data. And there are a large number of recruitment sites here in the broader Boston area where patients or any one of you, really, could go and volunteer to be part of this study. I just got a letter in the mail last week inviting me to go, and I was really excited to see that.

So all sorts of different data is being created as a result of these trends that I've been mentioning. And it ranges from unstructured data, like clinical notes, to imaging, lab tests, vital signs. Nowadays, what we used to think about just as clinical data now has started to really come to have a very tight tie to what we think about as biological data. So data from genomics and proteomics is starting to play a major role in both clinical research and clinical practice.

Of course, not everything that we traditionally think about healthcare data-- there are also some non-traditional views on health. So for example, social media is an interesting way of thinking through both psychiatric disorders, where many of us will post things on Facebook

and other places about our mental health, which give a lens on our mental health. Your phone, which is tracking your activity, will give us a view on how active we are. It might help us diagnose early the variety of conditions as well that I'll mention later.

So we have-- to this whole theme right now is about what's changed since the previous approaches at AI medicine. I've just talked about data, but data alone is not nearly enough. The other major change is that there has been decades' worth of work on standardizing health data.

So for example, when I mentioned to you that when you go to a doctor's office, and they send a bill, that bill is associated with a diagnosis. And that diagnosis is coded in a system called ICD-9 or ICD-10, which is a standardized system where, for many, not all, but many diseases, there is a corresponding code associated with it.

ICD-10, which was recently rolled out nationwide about a year ago is much more detailed than the previous coding system, includes some interesting categories. For example, bitten by a turtle has a code for it. Bitten by sea lion, struck by [INAUDIBLE]. So it's starting to get really detailed here, which has its benefits and its disadvantages when it comes to research using that data. But certainly, we can do more with detailed data than we could with less detailed data.

Laboratory test results are standardized using a system called LOINC, here in the United States. Every lab test order has an associated code for it. I just want to point out briefly that the values associated with those lab tests are less standardized.

Pharmacy, national drug codes should be very familiar to you. If you take any medication that you've been prescribed, and you look carefully, you'll see a number on it, and you see 0015347911, that number is unique to that medication. In fact, it's even unique to the brand of that medication. And there's an associated taxonomy with it. And so one can really understand in a very structured way what medications a patient is on and how those medications relate to one another.

A lot of medical data is found not in the structured form, but in free text, in notes written by doctors. And these notes have, often, lots of mentions of symptoms and conditions in them. And one can try to standardize those by mapping them to what's called a unified medical language system, which is an ontology with millions of different medical concepts in them.

So I'm not going to go too much more into these. They'll be the subject of much discussion in this semester, but particularly in the next two lectures by Pete. But I want to talk very briefly about what you can do once you have a standardized vocabulary.

So one thing you can do is you could build APIs, or Application Programming Interfaces, for now sending that data from place to place. And FHIR, F-H-I-R, is a new standard, which has widespread adoption now here in the United States for hospitals to provide data both for downstream clinical purposes but also directly to patients. And in this standard, it will use many of the vocabularies I mentioned to you in the previous slides to encode diagnoses, medications, allergies, problems, and even financial aspects that are relevant to the care of this patient.

And for those of you who have an Apple phone, for example, and if you open up a Apple Health Records, it makes use of this standard to receive data from over 50 different hospitals. And you should expect to see many competitors to them in the future, because of the fact that it's now an open standard.

Now other types of data, like the health insurance claims I mentioned earlier, is often encoded in a slightly different data model. One which my lab works quite a bit with is called OMOP, and it's being maintained by a nonprofit organization called the Observational Health Data Sciences Initiative Odyssey. And this common data model gives a standard way of taking data from an institution which might have its own intricacies and really mapping it to this common language, so that if you write a machine learning algorithm once, then that machine learning algorithm reads in data in this format, you can then apply it somewhere else very easily. And the portions of these standards really can't be understated, the importance for translating what we're doing in this class into clinical practice. And so we'll be returning to these things throughout the semester.

So we've talked about data. We've talked about standards. And the third wheel is breakthroughs in machine learning. And this should be no surprise to anyone in this room. All right, we've been seeing time and time again, over the last five years, benchmark after benchmark being improved upon and human performance beaten by state-of-the-art machine learning algorithms.

Here I'm just showing you a figure that I imagine many of you have seen, on the error rates on the image net competition for object recognition. The error rates in 2011 were 25%. And even

just a few years ago, it already surpassed human level to under 5%.

Now, the changes that have led to those advances in object recognition are going to have some parallels in healthcare, but only up to some point. For example, there was big data, large training sets that were critical for this. There were algorithmic advances, in particular convolutional neural networks, that played a huge role. And there was open source software that was created, things like TensorFlow and PyTorch, which allow a researcher or industry worker in one place to very, very quickly build upon successes from other researchers in other places and then release the code, so that one can really accelerate the rate of progress in this field.

Now, in terms of those algorithmic advances that have made a big difference, the ones that I would really like to point out because of their relevance to this course are learning with high dimensional features. So this was really the advances in the early 2000s, for example. And support vector machines and learning with L1 regularization as a type of sparsity.

And then more recently, in the last six years, on stochastic gradient descent, like methods for very rapidly solving these convex optimization problems, that will play a huge role in what we'll be doing in this course. In the last few years, there have been a huge amount of progress in unsupervised and semi-supervised learning algorithms. And as I'll tell you about much later, one of the major challenges in healthcare is that despite the fact that we have a large amount of data, we have very little labeled data. And so these semi-supervised learning algorithms are going to play a major role in being able to really take advantage of the data that we do have.

And then of course the modern deep learning algorithms. Convolutional neural networks, recurrent neural networks, and ways of trying to train them. So those played a major role in the advances in the tech industry. And to some extent, they'll play a major role in healthcare as well. And I'll point out a few examples of that in the rest of today's lecture.

So all of this coming together, the data availability, the advances in other fields of machine learning, and the huge amount of potential financial gain in healthcare and the potential social impact it could have has not gone unnoticed. And there's a huge amount of industry interested in this field. These are just some examples from names I think many of you are familiar with, like DeepMind Health and IBM Watson to startup companies like Bay Labs and PathAI, which is here in Boston, all of which are really trying to build the next generation of tools for healthcare, now based on machine learning algorithms.

There's been billions of dollars of funding in the recent quarters towards digital health efforts, with hundreds of different startups that are focused specifically on using artificial intelligence and healthcare. And there's the recognition that data is so essential to this process has led to an all-out purchasing effort to try to get as much of that data as you can.

So for example, IBM purchased a company called Merge, which made medical imaging software and thus had accumulated a large amount of medical imaging data for \$1 billion in 2015. They purchased Truven for \$2.6 billion in 2016. Flatiron Health, which is a company in New York City focused on oncology, was purchased for almost \$2 billion by Roche, a pharmaceutical company, just last year. And there's several more of these industry moves. Again, I'm just trying to get you thinking about what it really takes in this field, and getting access to data is actually a really important one, obviously.

So let's now move on to some examples of how machine learning will transform healthcare. To begin with, I want to really lay out the landscape here and define some language. There are a number of different players when it comes to the healthcare space. They're us, patients, consumers. They are the doctors that we go to, which you could think about as providers. But of course they're not just doctors, they're also nurses and community health workers and so on.

There are payers, which provide the-- where there is-- these edges are really showing relationships between the different players, so our consumers, we often, either from our job or directly from us, we will pay premiums for a health insurance company, to a health insurance company, and then that health insurance company is responsible for payments to the providers to provide services to us patients.

Now, here in the US, the payers are both commercial and governmental. So many of you will know companies like Cigna or Aetna or Blue Cross, which are commercial providers of healthcare, of health insurance, but there are also governmental ones. For example, the Veterans Health Administration runs one of the biggest health organizations in the United States, servicing our veterans from the department, people who have retired from the Department of Defense, which has the one of the second biggest health systems, the Defense Health Agency. And that is an organization where-- both of those organizations, where both the payer and the provider are really one.

The Center for Medicare and Medicaid Services here in the US provides health insurance for

all retirees in the United States. And also Medicaid, which is run at a state level, provides health insurance to a variety of individuals who would otherwise have difficulty purchasing or obtaining their own health insurance. And those are examples of state-run or federally run health insurance agencies. And then internationally, sometimes the lines are even more blurred. So of course in places like the United Kingdom, where you have a government-run health system, the National Health Service, you have the same system both paying for and providing the services.

Now, the reason why this is really important for us to think about already in lecture one is because what's so essential about this field is figuring out where the knob is that you can turn to try to improve healthcare. Where can we deploy machine learning algorithms within healthcare? So some algorithms are going to be better run by providers, others are going to be better run by payers, others are going to be directly provided to patients, and some all of the above.

We also have to think about industrial questions, in terms of what is it going to take to develop a new product. Who will pay for this product? Which is again an important question when it comes to deploying algorithms here. So I'll run through a couple of very high-level examples driven from my own work, focused on the provider space, and then I'll bump up to talk a bit more broadly.

So for the last seven or eight years, I've been doing a lot of work in collaboration with Beth Israel Deaconess Medical Center, across the river, with their emergency department. And the emergency department is a really interesting clinical setting, because you have a very short period of time from when a patient comes into the hospital to diagnose what's going on with them, to initiate therapy, and then to decide what to do next.

Do you keep them in the hospital? Do you send them home? If you-- for each one of those things, what should the most immediate actions be? And at least here in the US, we're always understaffed. So we've got limited resources and very critical decisions to make. So this is one example of a setting where algorithms that are running behind the scenes could potentially really help with some of the challenges I mentioned earlier.

So for example, one could imagine an algorithm which builds on techniques like what I mentioned to you for an internist one or quick medical reference, try to reason about what's going on with the patient based on the data that's available for the patient, the symptoms. But

the modern view of this shouldn't, of course, use binary indicators of each symptom, which have to be entered in manually, but rather all of these things should be automatically extracted from the electronic medical record or listed as necessary.

And then if one could reason about what's going on with a patient, we wouldn't necessarily want to use it for a diagnosis, although in some cases, you might use it for an earlier diagnosis. But it could also be used for a number of other more subtle interventions, for example, better triage to figure out which patients need to be seen first. Early detection of adverse events or recognition that there might be some unusual actions which might actually be medical errors that you want to surface now and draw attention to.

Now, you could also use this understanding of what's going on with a patient to change the way that clinicians interact with patient data. So for example, one can try to propagate best practices by surfacing clinical decision support, automatically triggering this clinical decision support for patients that you think it might be relevant for. And here's one example, where it says, the ED Dashboard, the Emergency Department Dashboard decision support algorithms have determined this patient may be eligible for the atrie cellulitis pathway. Cellulitis is often caused by infections. Please choose from one of the options. Enroll in the pathway, decline-- and if you decline, you must include a comment for the reviewers.

Now, if you clicked on enroll in the pathway, at that moment, machine learning disappears. Rather, there is a standardized process. It's an algorithm, but it's a deterministic algorithm, for how patients with cellulitis should be properly managed, diagnosed, and treated. That algorithm comes from best practices, comes from clinicians coming together, analyzing past data, understanding what would be good ways to treat patients of this type, and then formalizing that in a document.

The challenge is that there might be hundreds or even thousands of these best practices. And in an academic medical center, where you have patients coming-- where you have medical students or residents who are very quickly rotating through the system and thus may not be familiar with which are the most appropriate clinical guidelines to use for any one patient in this institution.

Or if you go to a rural site, where this academic nature of thinking through what the right clinical guidelines are is a little bit less of the mainstream, everyday activity, the question of which one to use when is very challenging. And so that's where the machine learning

algorithms can come in. By reasoning about what's going on with a patient, you might have a good guess of what might be appropriate for this patient, and you use that to automatically surface the right clinical decisions for a trigger.

Another example is by just trying to anticipate clinician needs. So for example, if you think that this patient might be coming in for a psychiatric condition, or maybe you recognize that the patient came in that triage and was complaining of chest pain, then there might be a psych order set, which includes laboratory test results that are relevant for psychiatric patients, or a chest pain order set, which includes both laboratory tests and interventions, like aspirin, that might be suggested.

Now, these are also examples where these order sets are not created by machine learning algorithms. Although that's something we could discuss later in the semester. Rather, they're standardized. But the goal of the machine learning algorithm is just to figure out which ones to show when directly to the clinicians. I'm showing you these examples to try to point out that diagnosis isn't the whole story. Thinking through what are the more subtle interventions we can do with machine learning and AI and healthcare is going to be really important to having the impact that it could have.

So other examples, now a bit more on the diagnosis style, are reducing the need for specialist consults. So you might have a patient come in, and it might be really quick to get the patient in front of an X-ray to do a chest X-ray, but then finding the radiologist to review that X-ray could take a lot of time. And in some places, radiologist consults could take days, depending on the urgency of the condition.

So this is an area where data is quite standardized. In fact, MIT just released last week a data set of 300,000 chest x-rays with associated labels on them. And one could try to ask the question of could we build machine learning algorithms using the convolutional neural network type techniques that we've seen play a big role in object recognition to try to understand what's going on with this patient. For example, in this case, the prediction is the patient has pneumonia, from this chest X-ray. And using those systems, it could help both reduce the load of radiology consults, and it could allow us to really translate these algorithms to settings which might be much more resource poor, for example, in developing nations.

Now, the same sorts of techniques can be used for other data modalities. So this is an example of data that could be obtained from an EKG. And from looking at this EKG, one can



try to predict, does the patient have a heart condition, such as an arrhythmia.

Now, these types of data used to just be obtained when you go to a doctor's office. But today, they're available to all of us. For example, in Apple's most recent watch that was released, it has a single-lead EKG built into it, which can attempt to predict if a patient has an arrhythmia or not.

And there are a lot of subtleties, of course, around what it took to get regulatory approval for that, which we'll be discussing later in the semester, and how one safely deploys such algorithms directly to consumers. And there, there are a variety of techniques that could be used. And in a few lectures, I'll talk to you about techniques from the '80s and '90s which were based on trying to signal processing, trying to detect where are the peaks of the signal, look at a distance between peaks. And more recently, because of the large wealth of data that is available, we've been using convolutional neural network-based approaches to try to understand this data and predict from it.

Yet another example from the ER really has to do with not how do we care for the patient today, but how do we get better data, which will then result in taking better care of the patient tomorrow. And so one example of that, which my group deployed at Beth Israel Deaconess, and it's still running there in the emergency department, has to do with getting higher quality chief complaints. The chief complaint is usually a very short, two or three word quantity, like left knee pain, rectal pain, right upper quadrant, RUQ, abdominal pain.

And it's just a very quick summary of why did the patient come into the ER today. And despite the fact that it's so few words, it plays a huge role in the care of a patient. If you look at the big screens in the ER, which summarize who are the patients and on what beds, they have the chief complaint next to it.

Chief complaints are used as criteria for enrolling patients in clinical trials. It's used as criteria for doing retrospective quality research to see how do we care for patients in a particular type. So it plays a very big role. But unfortunately, the data that we've been getting has been crap. And that's because it was free text, and it was sufficiently high dimensional that just attempting to standardize it with a big dropdown list, like you see over here, would have killed the clinical workflow. It would've taken way too much time for clinicians to try to find the relevant one. And so it just wouldn't have been used.

And that's where some very simple machine learning algorithms turned out to be really

valuable. So for example, we changed the workflow altogether. Rather than the chief complaint being the first thing that the triage nurse assigns when the patient comes in, it's the last thing. First, the nurse takes the vital signs, patient's temperature, heart rate, blood pressure, respiratory rate, and oxygen saturation. They talk to the patient. They write up a 10-word or 30-word note about what's going on with the patient.

Here it says, "69-year-old male patient with severe intermittent right upper quadrant pain. Began soon after eating. Also is a heavy drinker." So quite a bit of information in that.

We take that. We use a machine learning algorithm, a supervised machine learning algorithm in this case, to predict a set of chief complaints now drawn from a standardized ontology. We show the five most likely ones, and the clinician, in this case, a nurse, could just click one of them, and it would enter it into there.

We also allow the nurse to type in part of a chief complaint. But rather than just doing a text matching to find words that match what's being typed in, we do a contextual autocomplete. So we use our predictions to prioritize what's the most likely chief complaint that contains that sequence of characters.

And that way it's way faster to enter in the relevant information. And what we found is that over time, we got much higher quality data out. And again, this is something we'll be talking about in one of our lectures in this course.

So I just gave you an example, a few examples, of how machine learning and artificial intelligence will transform the provider space, but now I want to jump up a level and think through not how do we treat a patient today, but how do we think about the progression of a patient's chronic disease over a period of years. It could be 10 years, 20 years. And this question of how do we manage chronic disease is something which affects all aspects of the healthcare ecosystem. It'll be used by providers, payers, and also by patients themselves.

So consider a patient with chronic kidney disease. Chronic kidney disease, it typically only gets worse. So you might start with the patient being healthy and then have some increased risk. Eventually, they have some kidney damage. Over time, they reach kidney failure. And once they reach kidney failure, typically, they need dialysis or a kidney transplant.

But understanding when each of these things is going to happen for patients is actually really, really challenging. Right now, we have one way of trying to stage patients. The standard

approach is known as the EGFR. It's derived predominantly from the patient's creatinine, which is a blood test result, and their age.

And it gives you a number out. And from that number, you can get some sense of where the patient is in this trajectory. But it's really coarse grained, and it's not at all predictive about when the patient is going to progress to the next stage of the disease.

Now, other conditions, for example, some cancers, like I'll tell you about next, don't follow that linear trajectory. Rather, patients' conditions and the disease burden, which is what I'm showing you in the y-axis here, might get worse, better, worse again, better again, worse again, and so on, and of course is a function of the treatment for the patient and other things that are going on with them. And understanding what influences, how a patient's disease is going to progress, and when is that progression going to happen, could be enormously valuable for many of those different parts of the healthcare ecosystem.

So one concrete example of how that type of prediction could be used would be in a type of precision medicine. So returning back to the example that I mentioned in the very beginning of today's lecture of multiple myeloma, which I said my mother died of, there are a large number of existing treatments for multiple myeloma. And we don't really know which treatments work best for whom.

But imagine a day where we have algorithms that could take what you know about a patient at one point in time. That might include, for example, blood test results. It might include RNA seq, which gives you some sense of the gene expression for the patient, that in this case would be derived from a sample taken from the patient's bone marrow. You could take that data and try to predict what would happen to a patient under two different scenarios.

The blue scenario that I'm showing you here, if you give them treatment A, or this red scenario here, where you give them treatment B. And of course, treatment A and treatment B aren't just one-time treatments, but they're strategies. So they're repeated treatments across time, with some intervals.

And if your algorithm says that under treatment B, this is what's going to happen, then you might-- the clinician might think, OK. Treatment B is probably the way to go here. It's going to long-term control the patient's disease burden the best. And this is an example of a causal question. Because we want to know how do we cause a change in the patient's disease trajectory. And we can try to answer this now using data.

So for example, one of the data sets that's available for you to use in your course projects is from the Multiple Myeloma Research Foundation. It's an example of a disease registry, just like the disease registry I talked to you about earlier for rheumatoid arthritis. And it follows about 1,000 patients across time, patients who have multiple myeloma. What treatments they're getting, what their symptoms are, and at a couple of different stages, very detailed biological data about their cancer, in this case, RNA seq.

And one could attempt to use that data to learn models to make predictions like this. But such predictions are fraught with errors. And one of the things that Pete and I will be teaching in this course is that there's a very big difference between prediction and prediction for the purpose of making causal statements. And the way that you interpret the data that you have, when your goal is to do treatment suggestion or optimization, is going to be very different from what you were taught in your introductory machine learning algorithms class.

So other ways that we could try to treat and manage patients with chronic disease include early diagnosis. For example, patients with Alzheimer's disease, there's been some really interesting results just in the last few years, here. Or new modalities altogether. For example, liquid biopsies that are able to do early diagnosis of cancer, even without having to do a biopsy of the cancer tumor itself.

We can also think about how do we better track and measure chronic disease. So one example shown on the left here is from Dina Katabi's lab here at MIT and CSAIL, where they've developed a system called Emerald, which is using wireless signals, the same wireless signals that we have in this room today, to try to track patients. And they can actually see behind walls, which is quite impressive.

So using this for the signal, you could install what looks like just a regular wireless router in an elderly person's home, and you could detect if that elderly patient falls. And of course if the patient has fallen, and they're elderly, it might be very hard for them to get back up. They might have broken a hip, for example.

And one could then alert the caregivers, maybe if necessary, bring in emergency support. And that could have a long-term outcome for this patient which would really help them. So this is an example of what I mean by better tracking patients with chronic disease.

Another example comes from patients who have type 1 diabetes. Type 1 diabetes, as

opposed to type 2 diabetes, generally develops in patients at a very early age. Usually as children it's diagnosed. And one is typically managed by having an insulin pump, which is attached to a patient and can give injections of insulin on the fly, as necessary.

But there's a really challenging control problem there. If you give a patient too much insulin, you could kill them. If you give them too little insulin, you could really hurt them.

And how much insulin you give them is going to be a function of their activity. It's going to be a function of what food they're eating and various other factors. So this is a question which the control theory community has been thinking through for a number of years, and there are a number of sophisticated algorithms that are present in today's products, and I wouldn't be surprised if one or two people in the room today have one of these.

But it also presents a really interesting opportunity for machine learning. Because right now, we're not doing a very good job at predicting future glucose levels, which is essential to figure out how to regulate insulin. And if we had algorithms that could, for example, take a patient's phone, take a picture of the food that a patient is eating, have that automatically feed into an algorithm that predicts its caloric content and how quickly that'll be processed by the body. And then as a result of that, think about when, based on this patient's metabolic system, when should you start increasing insulin levels and by how much. That could have a huge impact in quality of life of these types of patients.

So finally, we've talked a lot about how do we manage healthcare, but equally important is about discovery. So the same data that we could use to try to change the way that algorithms are implemented could be used to think through what would be new treatments and make new discoveries about disease subtypes. So at one point later in the semester, we'll be talking about disease progression modeling, and we'll talk about how to use data-driven approaches to discover different subtypes of disease.

And on the left, here, I'm showing you an example of a really nice study from back in 2008 that used a k-means clustering algorithm to discover subtypes of asthma. One could also use machine learning to try to make discoveries about what proteins, for example, are important in regulating disease. How can we differentiate at a biological level which patients will progress quickly, which patients will respond to treatment. And that of course will then suggest new ways of-- new drug targets for new pharmaceutical efforts.

Another direction also studied here at MIT, by quite a few labs, actually, has to do with drug

creation or discovery. So one could use machine learning algorithms to try to predict what would a good antibody be for trying to bind with a particular target.

So that's all for my overview. And in the remaining 20 minutes, I'm going to tell you a little bit about what's unique about machine learning in healthcare, and then an overview of the class syllabus. And I do see that it says, replace lamp in six minutes, or power will turn off and go into standby mode.

**AUDIENCE:** We have that one [INAUDIBLE].

**DAVID SONTAG:** Ah, OK. Good. You're hired. If you didn't get into the class, talk to me afterwards. All right.

**AUDIENCE:** [INAUDIBLE].

**DAVID SONTAG:** [LAUGHS] We hope. So what's unique about machine learning healthcare? I gave you already some hints at this. So first, healthcare is ultimately, unfortunately, about life or death decisions. So we need robust algorithms that don't screw up.

A prime example of this, which I'll tell you a little bit more about towards the end of the semester is from a major software error that occurred something like 20, 30 years ago in a-- in an X-ray type of device, where an overwhelming amount of radiation was exposed to a patient just because of a software overflow problem, a bug. And of course that resulted in a number of patients dying.

So that was a software error from decades ago, where there was no machine learning in the loop. And as a result of that and similar types of disasters, including in the space industry and airplanes and so on, led to a whole area of research in computer science in formal methods and how do we design computer algorithms that can check that a piece of software would do what it's supposed to do and would not make-- and that there are no bugs in it. But now that we're going to start to bring data and machine learning algorithms into the picture, we are really suffering for lack of good tools for doing similar formal checking of our algorithms and their behavior.

And so this is going to be really important in the future decade, as machine learning gets deployed not just in settings like healthcare, but also in other settings of life and death, such as in autonomous driving. And it's something that we'll touch on throughout the semester. So for example, when one deploys machine learning algorithms, we need to be thinking about are

they safe, but also how do we check for safety long-term? What are checks and balances that we should put into the deployment of the algorithm to make sure that it's still working as it was intended?

We also need fair and accountable algorithms. Because increasingly, machine learning results are being used to drive resources in a healthcare setting. An example that I'll discuss in about a week and a half, when we talk about risk stratification, is that algorithms are being used by payers to risk stratify patients. For example, to figure out which patients are likely to be readmitted to the hospital in the next 30 days, or are likely to have undiagnosed diabetes, or are likely to progress quickly in their diabetes. And based on those predictions, they're doing a number of interventions.

For example, they might send nurses to the patient's home. They might offer their members access to a weight loss program. And each of these interventions has money associated to them. They have a cost. And so you can't do them for everyone.

And so one uses machine learning algorithms to prioritize who do you give those interventions to. But because health is so intimately tied to socioeconomic status, one can think about what happens if these algorithms are not fair. It could have really long-term implications for our society, and it's something that we're going to talk about later in the semester as well.

Now, I mentioned earlier that many of the questions that we need to study in the field don't have good label data. In cases where we know we want to predict, there's a supervised prediction problem, often we just don't have labels for that thing we want to predict.

But also, in many situations, we're not interested in just predicting something. We're interested in discovery. So for example, when I talk about disease subtyping or disease progression, it's much harder to quantify what you're looking for. And so unsupervised learning algorithms are going to be really important for what we do.

And finally, I already mentioned how many of the questions we want to answer are causal in nature, particularly when you want to think about treatment strategies. And so we'll have two lectures on causal inference, and we'll have two lectures on reinforcement learning, which is increasingly being used to learn treatment policies in healthcare.

So all of these different problems that we've talked about result in our having to rethink how do we do machine learning in this setting. For example, because driving labels for supervised

prediction is very hard, one has to think through how could we automatically build algorithms to do what's called electronic phenotyping to discover, to figure out automatically, what is the relevant labels for a set of patients that one could then attempt to predict in the future.

Because we often have very little data, for example, some rare diseases, there might only be a few hundred or a few thousand people in the nation that have that disease. Some common diseases present in very diverse ways and [INAUDIBLE] are very rare. Because of that, you have just a small number of patient samples that you could get, even if you had all of the data in the right place.

And so we need to think through how can we bring through-- how can we bring together domain knowledge. How can we bring together data from other areas-- will everyone look over here now-- from other areas, other diseases, in order to learn something that then we could refine for the foreground question of interest.

Finally, there is a ton of missing data in healthcare. So raise your hand if you've only been seeing your current primary care physician for less than four years. OK. Now, this was an easy guess, because all of you are students, and you probably don't live in Boston.

But here in the US, even after you graduate, you go out into the world, you have a job, and that job pays your health insurance. And you know what? Most of you are going to go into the tech industry, and most of you are going to switch jobs every four years. And so your health insurance is going to change every four years.

And unfortunately, data doesn't tend to follow people when you change providers or payers. And so what that means is for any one thing we might want to study, we tend to not have very good longitudinal data on those individuals, at least not here in the United States. That story is a little bit different in other places, like the UK or Israel, for example.

Moreover, we also have a very bad lens on that healthcare data. So even if you've been going to the same doctor for a while, we tend to only have data on you when something's been recorded. So if you went to a doctor, you had a lab test performed, we know the results of it. If you've never gotten your glucose tested, it's very hard, though not impossible, to figure out if you might be diabetic.

So thinking about how do we deal with the fact that there's a large amount of missing data, where that missing data has very different patterns across patients, and where there might be



a big difference between train and test distributions is going to be a major part of what we discuss in this course. And finally, the last example is censoring. I think I've said finally a few times.

So censoring, which we'll talk about in two weeks, is what happens when you have data only for small windows of time. So for example, you have a data set where your goal is to predict survival. You want to know how long until a person dies.

But a person-- you only have data on them up to January 2009, and they haven't yet died by January 2009. Then that individual is censored. You don't know what would have happened, you don't know when they died. So that doesn't mean you should throw away that data point. In fact, we'll talk about learning algorithms that can learn from censored data very effectively.

So there are a number of also logistical challenges to doing machine learning in healthcare. I talked about how having access to data is so important, but one of the reasons-- there are others-- for why getting large amounts of data in the public domain is challenging is because it's so sensitive. And removing identifiers, like name and social, from data which includes free text notes can be very challenging.

And as a result, when we do research here at MIT, typically, it takes us anywhere from a few months-- which has never happened-- to two years, which is the usual situation, to negotiate a data sharing agreement to get the health data to MIT to do research on. And of course then my students write code, which we're very happy to open source under MIT license, but that code is completely useless, because no one can reproduce their results on the same data because they don't have access to it. So that's a major challenge to this field.

Another challenge is about the difficulty in deploying machine learning algorithms due to the challenge of integration. So you build a good algorithm. You want to deploy it at your favorite hospital, but guess what? That hospital has Epic or Cerner or Athena or some other commercial electronic medical records system, and that electronic medical records system is not built for your algorithm to plug into. So there is a big gap, a large amount of difficulty to getting your algorithms into production systems, which we'll talk about as well during the semester.

So the goals that Pete and I have for you are as follows. We want you to get intuition for working with healthcare data. And so the next two lectures after today are going to focus on what healthcare is really like, and what is the healthcare data that's created by the practice of

healthcare like.

We want you to get intuition for how to formalize machine learning challenges as healthcare problems. And that formalization step is often the most tricky and something you'll spend a lot of time thinking through as part of your problem sets. Not all machine learning algorithms are equally useful. And so one theme that I'll return to throughout the semester is that despite the fact that deep learning is good for many speech recognition and computer vision problems, it actually isn't the best match to many problems in healthcare. And you'll explore that also as part of your problem sets, or at least one of them. And we want you to understand also the subtleties in robustly and safely deploying machine learning algorithms.

Now, more broadly, this is a young field. So for example, just recently, just about three years ago, was created the first conference on Machine Learning in Healthcare, by that name. And new publication venues are being created every single day by *Nature*, *Lancet*, and also machine learning journals, for publishing research on machine learning healthcare.

Because it's one of those issues we talked about, like access to data, not very good benchmarks, reproducibility has been a major challenge. And this is again something that the field is only now starting to really grapple with. And so as part of this course, oh so many of you are currently PhD students or will soon be PhD students, we're going to think through what are some of the challenges for the research field. What are some of the open problems that you might want to work on, either during your PhD or during your future career.