Welcome to Yale Cancer Answers with your host doctor Anees Chagpar.

Yale Cancer Answers features the latest information on cancer care by welcoming oncologists and specialists who are on the forefront of the battle to fight cancer. This week it’s a conversation about deep learning and cancer outcomes with Doctor Sanjay Aneja. Doctor Aneja is an assistant professor of therapeutic radiology at the Yale School of Medicine, where doctor Chagpar is a professor of surgical oncology.

Maybe we can start by you telling us a little bit about yourself and about your research?

Sure, I’ve been in New Haven since 2009. I was actually a Yale medical student and I stayed here for my residency and now on faculty.
Clinically, I am a radiation Oncologist and I primarily treat tumors of the central nervous system. So brain tumors as well and some prostate cancer. But I also run a large research group which is primarily focused on Applied Mathematics. My background is in applied mathematics, it’s always something that I was very interested in. And what we’re particularly interested in in my lab is looking at the utility of machine learning techniques, specifically deep learning in improving cancer outcomes and modeling cancer processes.

I was with you all the way up to applied mathematics and machine learning and deep learning and all of that sounds really deep. But can you break it down into simple terms for us? What exactly are you doing? That’s a good question, I think that we’re very interested in that we have such a large amount of healthcare data that’s currently available to us that’s been kind of cultivated over years of managing cancer patients, and what we’d like to do is develop methods to better model that data so...
that we can kind of use that information to better improve the health care outcomes for cancer patients moving forward. Machine learning is just one way in which we do that. Traditionally, the way in which we used to model cancer as a disease it’s always been a very difficult process. We would look at factors that physicians thought were important, and then we put them in a model and then we kind of look at an average and that had a couple different problems associated with it. One is that physicians aren’t really great at predicting what factors are actually associated with cancer. It’s a really complex disease. The second thing is that cancer is very difficult to model, and so using some of these techniques that we were developing a number of years ago, those techniques weren’t necessarily as effective. Machine learning is sort of an advanced form of modeling data. What it does is it takes in all types of data, so it doesn’t really require the physician to make any sort of choices.
about what type of data to include, and it allows us to model very complex processes like cancer, and there’s been a lot of effort within our group to show that machine learning methods are probably the best way in which we can model cancer outcomes, and so that’s what we’re particularly interested in.

Can you give us an example of how you did that in your lab?

The concept of taking large amounts of data variables of various sorts that physicians may or may not think are relevant at all, giving it to a computer and saying, hey, look at all of this data and see whether or not any of these things, or a combination of these things may actually predict a particular outcome. Am I on the right track?

Yeah, definitely.

I think the one benefit of machine learning, compared to more traditional techniques of modeling cancer data is that it allows us to look at all the various resources.

An example of a project that we’ve done in our lab is trying to model the outcomes for patients with early stage lung cancer. And so what we do in clinical practice is that we look at various different...
things when we’re trying to model early stage lung cancer patients. We look at certain demographic variables like your age and if they’re smoking and things of that nature, we also look at the images to see how big the tumor looks and whether it’s close to any structures that we’re worried about.

And then we also look at what are treatment plans are and so how well we can deliver radiation to treat those and those are kind of three different data sources of sorts that we use in clinic in order to determine whether or not a patient will have a good outcome or bad outcome. And currently the only models that we actually have to tell patients how their outcomes would be are using only demographic variables, and they’re not really using the pictures, and they’re not really using the treatment planning information, and the reason for that is because that data isn’t necessarily something that you can put into some of those traditional models.

What we did is we developed a deep learning machine learning algorithm. So it’s an algorithm that takes
the rawest form of the data from the electronic medical record, it pulls that demographic data from the electronic medical record, it takes every pixel from every picture of the tumor and analyzes those pixels in a very unique way. And then it also looks at every little part of our radiation treatment plan down to the pixel level. It kind of coalesces all that information and derives a personalized prediction, which we found was better than sort of getting an average based on just the demographic variables alone. So basically it’s taking all of this information, the clinical information that most clinicians would use, the image Ng that they also use, but that they can’t really put into a model because it’s hard to define, like I see a big tumor, it looks like it’s encasing some important vessels, but how do I really put that into a model and the treatment plan and can tell you kind of trying to be a clinician because a clinician will kind of look at that and have a good shift alt of.
This patient will do well. This patient won’t do so well and the computer can kind of give you that in a more quantitative way.

Yeah, I think that you’re touching upon one of the big advantages of these sorts of techniques. So one is that there an objective form, and so it’s not necessarily utilizing one physician’s experiences, or other physician experiences trying to use everyone’s collective experience of analyzing data in an objective way.

The other thing I think you’re kind of touching upon is this idea of we’re trying to mimic the same set of predictions that physicians make.

And that’s another reason why a lot of people are very interested in machine learning is because there is this component of artificial intelligence that can be kind of created when you are able to look at data sources without choosing which variables to evaluate. and so

In this project where you were looking at outcomes of early lung cancer and giving a machine the demographic data,
the smoking data, the imaging data, the treatment plan data, and you found that it was able to predict outcomes in terms of survival or in terms of recurrence. Yes, we found that it was able to predict recurrence in various different ways, survival, and when we compared it to maybe just using traditional methods or just one data stream, we found that it outperformed all of those different methods and so this idea of combining everything together is very very essential and we know it’s very intuitive for clinicians to realize you have to do that, but I think that it’s important to be able to do that in a mathematical way as well. Did it outperform the best guess of a clinician? So instead of looking at just traditional models, we know that clinicians sometimes bring their own experience and expertise to the equation. Did you compare the machine learning to clinicians best guess, how well patients would do or not do? Yeah, so one thing that we’ve done is we’ve done studies that are
looking at that, and similarly, what we’ve done is had multiple physicians do their best guess, and the first thing that’s important to know is that physicians don’t guess the same, and so there’s actually not a gold standard for a way in which a physician would actually evaluate a patient. And we found that it performed at least as well as an experienced clinician, and better than maybe less experienced clinicians. In a sense, you’re recreating with this machine learning the predictive ability of an experienced clinician. So how is that now being utilized? Or is it being utilized in the clinic? I think that one of the benefits of the platform that we’ve developed is that it doesn’t actually require us to pull data and put it into a calculator of sorts, which is a lot of what we see with a lot of predictive things in cancer. And so what we’re interested in right now is trying to connect ours to the electronic medical record. We’ve developed an iPhone application which allows us to basically put
0:08:49.656 -> 0:08:51.76 in the medical record number of a patient,
0:08:51.76 -> 0:08:54.1 and then it allows us to pull the data
0:08:54.1 -> 0:08:56.165 natively and then it allows us to kind
0:08:56.165 -> 0:08:58.357 of develop that prediction in the clinic,
0:08:58.36 -> 0:08:59.75 and that’s the next step
0:08:59.75 -> 0:09:01.38 of what we’re trying to do.
0:09:01.38 -> 0:09:03.193 But I think the other thing that’s
0:09:03.193 -> 0:09:04.317 really important whenever we’re
0:09:04.317 -> 0:09:05.827 thinking about these machine learning
0:09:05.827 -> 0:09:07.574 algorithms is because they are so
0:09:07.574 -> 0:09:08.804 good at modeling healthcare data,
0:09:08.81 -> 0:09:10.52 they tend to actually model it
0:09:10.52 -> 0:09:13.04 too well, and so that’s what we call overfit.
0:09:13.04 -> 0:09:13.769 The data set,
0:09:13.769 -> 0:09:15.859 and so they sort of are very good
0:09:15.859 -> 0:09:17.797 at modeling Yale data for example,
0:09:17.8 -> 0:09:19.704 but they might not be so good
0:09:19.704 -> 0:09:20.88 at modeling data from
0:09:20.88 -> 0:09:21.804 I don’t know Chicago,
0:09:21.804 -> 0:09:23.527 and what we’re trying to do
0:09:23.527 -> 0:09:24.962 also is something called external
0:09:24.962 -> 0:09:26.787 validation where we send our model
0:09:26.787 -> 0:09:28.031 to different cancer
0:09:28.031 -> 0:09:29.84 centers across the country and say,
0:09:30.12 -> 0:09:32.36 don’t even tell us what the outcomes are,
0:09:32.36 -> 0:09:34.32 we will tell you what our models are predicting,
0:09:34.32 -> 0:09:36.301 and then you tell us how good
0:09:36.301 -> 0:09:37.679 our model did so far.
0:09:37.68 -> 0:09:38.824 Our collaboration with Jefferson
0:09:38.824 -> 0:09:40.254 in Philadelphia has shown that
0:09:40.254 -> 0:09:41.67 the model is very productive.
And it’s maintaining that same performance, but it’s important to test these models before we actually put them into clinical practice. How well something can predict is based on how well we learned and on the learning set that it had to work with. It makes sense, however, that the data that it was getting was objective data. It wasn’t getting data that may have had a lot of factors that were subjective, right? You were looking at imaging. Well, the image is what it is and if you give that image to two different radiologists, they both should say roughly the same thing, maybe not exactly, but roughly.

Jefferson images are going to be very much like Yale images and so that may account for that close correlation between the two datasets. But the next question is OK, let’s suppose that the model after you test it and I don’t want to minimize the utility of making sure it’s externally generalizable. Is it even being used here at Yale, where it was developed and it does well in terms of predicting
outcomes as well as an experienced clinician?
Is that being used in the clinic?
Are you putting in this data
you've got now,
this iPhone application that can
pull in this data into this model?
The model can do its magic and tell you,
this is the recurrence rate.
This is the survival rate.
Are you using that in the clinic,
and if so how?
I think that we’re in the process
of developing the application.
One of the big hurdles and with health
care in general is the ability to
actually get access to the electronic
medical record in a way in which
you can make an application kind
of seamlessly integrate into it.
And so it’s somewhat difficult
for us to do that.
We’re working with
a software engineering firm
to actually help us with
beyond the scope
of what our lab does typically.
And so that’s where we’re
right now with respect to actually
integrating into clinical practice.
I think that right now what we
have is we have an ability to
kind of look back on patients, and if there was a patient, for example, who wanted to have a prediction, we could actually generate that. But we cannot do it in the electronic medical record. As of right now, and so when you think about the potential utility of this, where do you see it going? So I think that one thing that I think is very important is as we’re kind of developing so many different genres of treatment for cancer patients, there’s this increasing need for us to develop methods to risk stratify them and identify the highest risk. Patients who maybe would be benefiting from more aggressive treatment, more aggressive followup, I think we’ve found with some types of cancers that maybe we’ve been a little bit too aggressive in our follow-up or too aggressive with our therapy and to risk stratify which of those patients would be most useful for certain intervention versus another one is something that I think are our algorithm
Our platform is very useful, especially for early stage lung cancer patients. There's a clinical trial evaluating whether or not patients should get radiation and additional treatment. The idea is that potentially additional immunotherapy, for example, would be helpful for some patients. A large amount may not need any additional treatment, while a large amount of patients will need it after treatment. We can identify high-risk patients versus low-risk patients, which can help tailor our treatments and better understand personalized care.
What would be interesting though, is to really look at how do patients do without any therapy? How do patients do with therapy X versus therapy Y? And how can we really personalize therapies given the data that we have?

We are going to take a short break for a medical minute, but when we come back we will answer those questions.

So stay tuned to learn more about deep learning in cancer outcomes with my guest doctor Sanjay Aneja.

Support for Yale Cancer Answers comes from AstraZeneca, a biopharmaceutical business that is pushing the boundaries of science to deliver new cancer medicines. More information at astrazeneca-us.com.

This is a medical minute about pancreatic cancer, which represents 3% of all cancers in the US and about 7% of cancer deaths.

Clinical trials are currently being offered at federally designated comprehensive cancer centers for the treatment of advanced stage and metastatic pancreatic cancer.
cancer using chemotherapy and other novel therapies like FOLFIRINOX a combination of five different chemotherapies is the latest advances in the treatment of metastatic pancreatic cancer, and research continues in centers around the world looking into targeted therapies. And a recently discovered marker hENT 1. This has been a medical minute brought to you as a public service by Yale Cancer Center. More information is available at yalecancercenter.org. You’re listening to Connecticut public radio. Welcome back to Yale Cancer Answers. This is doctor Anees Chagpar and I’m joined tonight by my guest doctor Sanjay Aneja we’re discussing deep learning in cancer outcomes and right before the break, doctor Aneja was telling us about how he and his lab have really used machine learning. That is to say, Applied Mathematics and complicated computational models to really take in lots and lots of data that clinicians using their usual clinical guess to
predict outcomes for cancer patients. And Sanjay, the example that you gave us in early lung cancer where you said, the demographics we took the imaging data, but then we took the treatment plan and we use that and we ask the computer to look at these things down to the pixel level and then predict outcomes. And it was very good at predicting those outcomes as good as an experienced clinician, and that’s great, but my question is what happens if you don’t? How do you get rid of the treatment part of that and predict outcomes without treatment to kind of get at the idea of, are we over treating some patients? Because if the computer doesn’t have that, if all patients are treated, and that’s the basis on which it learned, how do you take out one part of that model? That’s a good question, and it’s something that I think is not impossible. So I think that if we remove the treatment piece of it and just look at the images and the demographic data,
basically pretreatment information, we find the model is actually quite predictive as well. It just improves significantly if we know exactly what types of treatments we provided for the patients, and so another example of a study that we've done which only uses pretreatment imaging has been evaluating lymph nodes in head and neck cancer patients. We were attempting to look at which lymph nodes we saw on CT imaging actually had the presence of cancer and we wanted to identify that so that maybe you know what we could do is more tailor the therapy for head and neck cancer patients. Oftentimes with head and neck cancer patients when we think that their lymph nodes don’t have cancer, we have them undergo surgery and then we find those lymph nodes have cancer. They have to get radiation and chemotherapy altogether. And so if we were able to identify the patients ahead of time that have cancer in the lymph nodes than what they would have instead is just chemotherapy and radiation.
They save themselves some surgery and so that's an example of when we've used pretreatment imaging to sort of reduce potential extra care or care that maybe would not be necessary. Do we have datasets with patients who were treated in different ways so that we can predict given pretreatment data, if you got treatment a, you will do this well, if you got treatment b you will do that well and if you got treatment c you will do this well. And if you got no treatment you would do just as well as any of the above. Yeah, so that actually kind of touches upon something that we're really actively exploring and NOTE Confidence: 0.904884934425354 something that we're very excited about. So one thing that we've kind of indicated, that these machine learning algorithms, these deep learning algorithms are extremely good at analyzing pictures, and so one thing that we've looked at is this idea of what we call digital Twins. So, based on your pretreatment imaging, nothing else no demographics,
just what your tumor looks like, if we could find your digital twin or someone who’s tumor looks exactly like yours or digital family, which is maybe a group of five people that are like that, we can use deep learning to do that. And then what we can do is we can see, OK among your digital twins, or your digital family who got one type of treatment, this is what their outcome was and among your digital family who got another type of treatment, this is what their outcome was. And then they can make a more informed decision about what they would actually want to do. We talk a lot on this show about tumor heterogeneity and different kinds of cancer and tumor biology, they seem to affect outcomes, so how is it that you can use that data without knowing all of the other things? I think for one, there’s a lot of evidence to suggest that deep learning algorithms can actually predict changes
in tumors up to the genomic level, so genetic mutations and tumor is based off of the pictures because we have to appreciate that their really evaluating every tumor at a very, very small level, every little pixel and each pixel has a variety of different intensities, and so they’re really looking at the data in a close level. So in lung cancer as well as in brain tumors, an also in some lymphomas there’s been evidence to suggest that deep learning algorithms on the diagnostic images can predict genomic changes. So like driver gene mutations that would actually maybe presumably need sequencing information for and so then that suggests that the pictures actually have a lot more information than we think, but I do think that you’re kind of right in one way that maybe it’s not just the pictures that tell the whole story. But the idea is that people whose tumors look similar, they likely have similar genomic backgrounds. I’m still puzzled by this whole concept because we think about a CT scan or a mammogram, we think about a CT scan or a mammogram, or an MRI and each of these has its own pitfalls. There are false positives on these images,
there are false negatives on these images, so it kind of makes me a little wary to put so much faith just in the images. How do you explain that?

I guess one of the limitations or one of the important caveats to any sort of machine learning project is that your outcomes are only as good as your data, and so if we have a lot of false positives in our data set that we have not addressed and we haven’t identified, then I think that it’s really important that we understand that the machine learning algorithm will learn those same errors. So similarly any sort of biases that we have, maybe we have a bias towards over imaging or over diagnosing something on an image, those similar biases will be promulgated through our machine learning algorithms. It’s actually somewhat of an interesting topic in the context of machine learning outside of healthcare is that we find that a lot of machine learning algorithms they mimic the same biases and discriminatory abilities that people have in regular practice, and so a lot of algorithms that have been used in law enforcement, we find that are actually maybe
promulgating some of the parts of our law enforcement that we don’t want. One way to kind of get around that one would think is to use more data, not just the images, but get down to the genomic level, do the biopsy. We’ve got biopsies on most tumors before we ever treat them, so look at the pathologic information, look at the genomic information. We can get a lot of sequencing data these days. And speaking of which, it’s really hard to understand what all of these different mutations are when we think about whole exome sequencing. I mean, I would think that machine learning might have a role to play there too. Certainly, I think that machine learning in general is probably one of the more common approaches to evaluate genomic data. Now, because the genome is so complex and it’s so difficult for us to kind of understand that the machine learning algorithms are maybe the most common ways in which we analyze that sort of information now, specifically with respect to deep learning, which is what our lab is,
particularly just specific types of machine learning methods. It’s somewhat difficult to evaluate genomic information with that data, and the reason for that is because we don’t actually have a huge data set right now at our disposal of genomic information, because in order to do some of these analysis, we need upwards of 1000 patients, and so it’s difficult to get 1000 patients with tumors, images, and whole exome sequencing, but it’s possible, and I would venture that a place like Yale is the place that would have that ability to do that or some of these cooperative groups, right? For our listeners, there are clinical trials that happen all across the country, sometimes all across the world with cooperative groups. These groups of clinicians, physicians who are all putting their patients on exactly the same clinical trial and taking their data, putting it in a central repository where all of that can be studied. Sanjay, I would think that
0:23:58.242 –> 0:24:00.539 would be an ideal place for you
0:24:00.539 –> 0:24:02.85 to get that data.
0:24:02.85 –> 0:24:04.873 Yes, and one effort of our research
group is actually sort of engaging
with the cooperative groups.
0:24:08.26 –> 0:24:10.084 There’s two that we’ve
begun engaging with, the NRG
0:24:11.76 –> 0:24:13.824 which is a large group that has a
lot of radiation data as well as the
Southwest Oncology Group also
0:24:18.839 known as SWOG in order to sort of develop
an infrastructure within the
organization to evaluate machine
learning techniques and utilized
machine learning techniques.
Because a lot of what
we’ve designed, these clinical
trials and these repositories is that
their infrastructure wasn’t made for
these sorts of analysis because they
weren’t necessarily thinking that
this is something that was going to
come on the horizon.
And so one thing that we’re working
right now is with SWOG and NRG
to develop that sort of infrastructure.
The first process of that is
developing something that allows us
to get the imaging data very easily.
Images are sort of an easy
method for us to evaluate machine learning methods because one, it’s been shown to be the most effective in image analysis across various industries, and secondly imaging in healthcare has a standardized data format. It’s a common data model, so there’s no difficulty about well, so and so in California stores their data one way, and then we store it a different way, etc. One thing that you mentioned, which I still have to go back to, is you said that you’re interested in deep learning, which is a type of machine learning that is particularly well suited to imaging. Tell us the difference between deep learning and machine learning. Yeah, that’s a good question, and the words, artificial intelligence, machine learning, deep learning are often used interchangeably, but they have distinct meanings.
discipline of various types of mathematical techniques to model data. Deep learning is just one of those techniques. Now the difference between deep learning and other traditional machine learning techniques is that other machine learning techniques require, you know inputs that are called features and so they can only handle data that comes in a featured format. So sort of predictor variables that you’re interested in demographic variables or variables from the electronic medical record. Deep learning is particularly unique in that it doesn’t actually require data at all from a human. It doesn’t require any sort of human interaction. It can learn those features on its own as long as it has access to what they called the sensor. So where the data is generated and as data is being generated in real time, deep learning algorithms can analyze it, identify those features that are very important, so those predictors are important and then create predictions. How exactly does this happen?
Somebody’s gotta program this thing right?
Yes, it is programmed, typically in Python and so the way that the process works for developing a deep learning algorithm is first you have a set of training data. And the associated labels to that data. So you already have data with outcomes that you know, and that’s the process you’re trying to predict and then what you do is you design your deep learning algorithm using a complex series of what they call neural networks and we kind of train the algorithm by looking at each of those training data set labels to identify sort of patterns in the data. It takes a significant amount of time and a huge amount of computational resources in order to do that. So deep learning algorithms that we developed in our lab oftentimes take weeks to train. And so we just let it run all week and let it run every single time and every piece of information that it gets looks at the outcome and it tries to learn.
0:27:46.228 –> 0:27:48.26 a little bit more from it.
0:27:48.26 –> 0:27:50.647 And so presumably you
design this algorithm, you let it run,
0:27:52.962 –> 0:27:56.03 the machine tries to learn what it can to
try and improve its prediction each time,
0:27:58.49 –> 0:28:01.56 and then you test it on a separate set.
0:28:01.56 –> 0:28:04.666 Yes, exactly and in order for
0:28:04.666 –> 0:28:07.898 this to be utilized because
0:28:07.9 –> 0:28:11.127 I’m thinking about how this can be
0:28:11.127 –> 0:28:14.582 utilized in the clinic in terms of you
0:28:14.582 –> 0:28:18.41 know one day you may walk into a clinic,
0:28:18.41 –> 0:28:21.476 have your CT scan, have your biopsy,
0:28:21.48 –> 0:28:24.192 fill out some paperwork on your
demographics and I can imagine a time
0:28:27.349 –> 0:28:30.808 when all of that information is put into
0:28:30.808 –> 0:28:33.888 a computer or stored in the computer.
0:28:33.89 –> 0:28:36.902 This algorithm runs in the background
0:28:36.902 –> 0:28:40.309 and spits out to the clinician,
0:28:40.31 –> 0:28:42.99 this patient’s prognosis is X.
0:28:42.99 –> 0:28:47.814 The ideal treatment out of A, B&C is B
0:28:47.82 –> 0:28:49.878 and that kind of thing.
0:28:50.17 –> 0:28:51.35 I could see that happening
0:28:51.35 –> 0:28:52.916 in the future. I think we’re
0:28:52.916 –> 0:28:55.004 a little bit far away from
0:28:55.004 –> 0:28:56.639 complete automation in that way,
0:28:56.64 –> 0:28:58.464 and partly because I think that
0:28:58.464 –> 0:29:01.078 we don’t have a good hold on
0:29:01.078 –> 0:29:03.371 the data that we think is the most
0:29:03.371 –> 0:29:05.457 important and we don’t have a good
0:29:05.46 –> 0:29:07.218 way of storing all that information,
0:29:07.22 –> 0:29:09.695 but I think that it’s not something that I
would be surprised we’re doing in 5-10 years.

Doctor Sanjay Aneja is an assistant professor of therapeutic radiology at the Yale School of Medicine.

If you have questions, the address is canceranswers@yale.edu and past editions of the program are available in audio and written form at Yalecancercenter.org.

We hope you’ll join us next week to learn more about the fight against cancer here on Connecticut public radio.