MCIDT Innovations in Practice: Introduction and Application of Machine Learning in Infectious Disease

Podcast Transcript

Sriram Chandrasekaran, PhD

November 7, 2024

PUBLIC HEALTH PREPARED
A Program of the Michigan Center for Infectious Disease Threats & Pandemic Preparedness

Public Health Prepared
University of Michigan
School of Public Health
1415 Washington Heights
Ann Arbor, MI 48109-2029

PHPrepared@umich.edu
734-615-0144
PublicHealthPrepared.org

0:00:11.1 Storee Harris: Welcome to Prepared. Set. Go! A podcast of Public Health Prepared. Public Health Prepared is a workforce development branch of the Michigan Center for Infectious Disease Threats and Pandemic Preparedness, or MCIDT Initiative, which is housed that and funded by the University of Michigan. We hope this podcast will better equip the public health workforce to handle ongoing and future health crises. Thank you for tuning into our episode.

0:00:36.6 SH: Today, we'll be talking about MCIDT Innovations and Practice Introduction and Application of Machine Learning and Infectious Disease. I'm your host, Storee Harris-Stubblefield.

0:00:46.3 SH: Today we have Dr. Sriram Chandrasekaran. Dr. Chandrasekaran is an Associate Professor and Associate Chair of Research in Biomedical Engineering at the University of Michigan-Ann Arbor. He leads the Systems Biology and Drug Discovery lab. He received his PhD in biophysics from the University of Illinois at Urbana-Champaign and worked as a Harvard Junior Fellow at Harvard University and MIT.

0:01:13.9 SH: He has developed over 10 systems biology methods for drug discovery and bioengineering. A key focus of his lab is the development of mechanistic AI methods and the use of engineering models and machine learning. He teaches a course called AI in BME that introduces students to AI algorithms and their application in biomedical engineering. He is a recipient of several awards, including the Howard Hughes Medical Institute, international Predoctoral Fellowship, the MIT Technology Reviews Top Innovators Under 35 Award, the EBS Teaching Award, and the Machine Learning and the Chemical Sciences Award from the Camille and Henry Dreyfus Foundation. Welcome to the podcast. I'm excited to speak with you today.

0:02:00.2 Sriram Chandrasekaran: Thank you, Storee. Excited to be here. Thanks for having me.

0:02:03.9 SH: Sure. So as we get started, let's just start talking about machine learning. So what is machine learning and how does it differ from AI?

0:02:13.3 SC: I would say they're related terms. So AI is something, AI means artificial intelligence. So it means anything that performs intelligent activities like an Alexa or a robot, you know anything, can be considered an AI. Machine learning is a subtype of AI where it learns patterns from data like it could be mathematical statistical patterns. For example, Gmail spam filter or a Netflix movie recommendation. Those are all examples of good machine learning algorithms where algorithm trains on your past movies, or like spams and then it predicts new TV shows or products that you would like. So it's trained on a specific data and predicts a specific thing. So that's typically what people call as machine learning. And more recently, there's more element of what's called a general purpose AI or GenAI like ChatGPT that is trained on variety of data and it makes... Does a variety of tasks. It can write poetry or create images. So now that's one of the latest generations of AI.

0:03:19.5 SH: Thank you so much for differentiating that for us. I think they're used interchangeably, and so there are even similar contexts, so it's good to know the difference between the two. Can you start telling us about a little bit about your background and what drew you to the intersection of machine learning and infectious disease research?

0:03:36.9 SC: Sure. Yeah. So, believe it or not, so my inspiration for learning AI and machine learning was actually on studying infectious diseases. So my PhD was in biophysics. I was developing theoretical models of biological systems. But then when I start thinking about how to tackle infectious disease, how to design drugs against different pathogens, I realized that given the diversity of drugs, diversity of pathogen strains, it's not possible to solve this problem experimentally or theoretically. You need machine learning tools to analyze this complex data and also come up with more meaningful predictions in a really quick time. So that's what drew me to a machine learning in the first place.

0:04:21.8 SC: And one of the first methods of using machine learning that I developed called Indigo, it tries to match drugs, drug combinations to specific pathogens. And I know it is a simple machine learning tool, but it's... I found that I was really surprised by its accuracy, was able to generalize really well, able to predict the thoroughness of different drugs against different pathogens. So that got

me really excited about machine learning a decade ago, and I'm still using that now because I really can see the potential of AI in infectious disease.

0:04:56.0 SH: Thank you for telling us a little bit about your background and being interested in infectious disease. I'm sure our audience is excited to hear that. I just wanna pivot a little bit more to the present. So how are you using machine learning in infectious disease research today?

0:05:10.7 SC: Yeah, so, I would say that we are like trying to tackle different problems. I guess the two main problems that we are trying to tackle are like, trying to understand the inter person variation, person to person variation in infection. For example, how does like tuberculosis or cold infection change from each person to another because you see a lot of variability. So using AI we can integrate data from the patient, but also the pathogen, like what strains and what its properties are, what mutations it has. And so by combining this diverse data, we can make more personalized predictions. And that's our goal, is we wanna combine different data sets and come up with medicine or personalized strategies for tackling infections.

0:06:01.9 SH: With making it so personalized, what type of challenges do you have when it comes to doing that?

0:06:06.8 SC: I guess in terms of machine learning challenges, I would say that making it more generalizable is a big challenge, for example, you may be trained on one type of data or type of pathogen, kind of generalized to others or other patient populations. So that's a big challenge with machine learning. And so we've been developing new types of AI to solve those problems. And I know there are also other problems like when you make predictions, you wanna be very specific to the pathogen of interest and you don't want to affect the native microbiome too much. So we've been trying to come up with algorithms that are very specific, identify drugs specific to a pathogen of interest without collateral damage to native microbiota. So there are two reasons for this, one is that it reduces the chance of resistance developing because you're targeting a very specific strain, and also obviously reduces risk of side effects. And so that's another area we're working on.

0:07:06.2 SH: That makes sense. Thank you for sharing that. Thinking about the challenges, what are some limitations of the current machine learning approaches in this field and how are you and your team working to overcome those?

0:07:20.0 SC: Yeah, sure. Yeah. Like I said, one is making AI more generalizable. And for that, we have been developing what's called transfer learning-based methods that usually generalize well beyond the data. So it's a type of machine learning where you are... Train on one model organism and then try to extrapolate predictions to related strains. So if there's a new pathogen coming in, pathogen strain coming in, we already have so much data on certain model organisms. We can try to extrapolate, identify drugs or predict its properties based on that information. So that is one area.

0:07:52.0 SC: Another one is also, or in terms of drug discovery, we are really interested in trying to find drugs that are potent against the pathogen, but non-toxic to human cells, so have reduced side effects. For many antibiotics, there's a lot of side effects of the kidney, liver. So we are trying to develop AI tools that not only predict just the potency, in that case if you just say, predict one thing, then you might predict bleach as the top predictor like for killing all the bugs. But you don't wanna do that. So if you wanna come up with something that's highly specific to the pathogen without collateral damage. And so they've been developing models that incorporate properties of human cells and the pathogen, and then try to identify those that have the most potency without side effects.

0:08:39.4 SH: Thank you. I just want to... I wanna make sure I didn't miss this, but I know that you are working on tuberculosis. Is there other infections that these are of interest right now or just tuberculosis?

0:08:50.3 SC: Oh, yeah. Thanks for asking that. Yeah. So yeah, we do focus on gram-negative infections and tuberculosis. Tuberculosis has been a big interest of mine for a long time, but we've also been slowly branching out to other gram-negative pathogenic, like E. Coli, Acinetobacter, and a few others.

0:09:04.9 SH: Awesome. Thank you. So in thinking a little bit more on the AI machine learning side, what are some ethical considerations when using machine

learning and things such as bias or the environmental impacts of running these machines?

0:09:22.8 SC: Yeah, that's a great question. So yeah, that's something we always think about is what are the inherent biases in our algorithm? And I would say their... Our main source of bias is the data that we feed into these algorithms. And so there are certain populations that are more well-studied, more pathogens, diseases that are more studied than others because there are more funding for certain things than others. And so that, once again, has ramifications on the algorithms we build. So the models that are being trained on patient populations or pathogens that have a lot of data would do better than others.

0:09:58.7 SC: And so from our end, what we've been trying to do is utilize methods like transfer learning so that we can leverage this data from one population and apply it to a new population with slightly just small tweaks in the model. So just to explain this better, so a really cool example of transfer learning is like a lot of cat videos on YouTube, right? So you can really build really good machine learning models to predict... Identify cats in videos, right? So, but then if you wanna apply the real world challenge of, let's say, tiger conservation, like you want to identify what tigers are there, there isn't much data on tigers as much as cats. So what people are trying to do now is take the machine learning model that was built on identifying cats and then just with an additional training data on tigers. This model can now be applied to predict, identify different tigers, right? So this wasn't possible before because there wasn't enough data on tigers.

0:10:51.8 SC: So that's something similar to what we're trying to do now is there might be certain pathogens, let's say E. Coli, for example, has a ton of data, right? So we can build machine learning models really well for an E. Coli. But then for diseases like tuberculosis or others where there isn't much data, we can try to tweak the model that was built on E. Coli with a little bit of additional data on tuberculosis and make that a TB-specific model. And so that's the idea behind transfer learning, is that many conditions or like disease or patient populations might not have enough data to build the model, but we can leverage data that's already out there that's more abundant and try to incorporate and tweak that and make it specific to the pathogen or patient population of interest. So that's one area we are trying to do.

0:11:36.0 SC: A second area that we are also really interested and also that goes directly to this issue of bias is that we are trying to make these machine learning models more transparent or mechanistic. You can imagine that when we are trying to incorporate different datasets, for example, for tuberculosis, a study that we did recently, we had to incorporate data from like patient chest X-rays, pathogen genomics, patient's demographics and others. So we have like hundreds of features. And so it's really hard to then understand what the machine learning algorithm is predicting or what it's using. And so we've been developing what's called a mechanistic or transparent AI that helps you explain, that explains how the predictions are made.

0:12:18.7 SC: So then for each patient, it will tell you, oh, this patient would do well because they have these properties. And so then we can go back and see why the AI algorithm made a certain recommendation for a certain patient and we can then test if there were biases in there. Was there like certain identifying features that were used that shouldn't have been used or there were biases in the data that we can easily identify? So those are two main ways.

0:12:44.3 SC: On the engineering side, we are trying to do this. One is transfer learning. Another is making AI more transparent or mechanistic.

0:12:54.4 SH: Thank you so much for speaking to those issues and bringing that to the forefront. In thinking about the achievements that you have made, what are some exciting achievements or advancements or emerging trends in machine learning that you think will shape the future of infectious disease research?

0:13:11.3 SC: Oh yeah, sure thing. This is an exciting time for AI machine learning, especially for infection too. So there's a lot of data sets that are now available partly due to the COVID pandemic and others. So there's a lot of open source data. And there's also great developments in AI, like for example, Generative AI tools are being applied to come up with completely new chemical entities or drugs. And so these might be drugs that evolution has not seen before, so maybe it takes longer to develop resistance. Those are all really exciting developments in the AI machine learning field, I think. And also there's also these general purpose tools that you can

take and repurpose for different applications. So those are two main exciting things I'm looking forward to.

0:14:01.0 SH: Awesome. I'm looking forward to it as well. I'm very interested in this topic. As I listen to you talk, I'm like, "Wow, this is really exciting." So AI is a hot topic, especially in public health, also in emergency preparedness and infectious disease. If you had to pitch AI to someone who was hesitant or had a lot of questions, what would you say is the main value add?

0:14:22.0 SC: Definitely. I think everyone talks about AI being fast and efficient, so on. For me, as a scientist, I feel like AI has no barriers. Like it truly gives you sometimes out-of-the-box predictions that you would've normally not thought about. To give you an example like that, this was once again, several years ago when we were originally working on AI models, the Indigo model that we developed for tuberculosis, it predicted some neurological drugs as being effective against tuberculosis. And we thought, hey, that's not possible. It didn't make sense, but then we went to the lab and tested those predictions and they turned out to be true. We had one antipsychotic drug that was highly synergistic with the tuberculosis drug, which was like, oh, super cool. And that shows that no, AI doesn't have these preconceived notions of what a drug should look like or what an antibiotic should look like. And so it can truly reduce the cost of drug development or identify really novel candidates. Maybe they're already out there and we have just not been looking at the right place. So that's what I'm really excited about AI.

0:15:24.8 SH: That's really interesting. Thank you for sharing that. We have a wide audience of people, so I would like to know, can you share any advice for young researchers or students that are interested in this interdisciplinary field?

0:15:39.0 SC: Oh yeah. Sure thing. This is an exciting time, as I said, and there's a lot of data that you don't even need to have, like work in a microbiology lab to start contributing to infectious disease. And there's a lot of these open source initiatives from NIH, the UK Biobank, now the TB Portals data, those are all publicly available. And so I would recommend students to download or access latest AI tools, which are also freely available and apply it to these emerging data so you can get your feet wet on what AI could do. I think it's a great combination of having the right data and the algorithm and there's a big need for infectious disease. So I think there's a

perfect synergy of three different things. And yeah, I would definitely encourage young people to look into all these data and start playing with it.

0:16:36.5 SH: Awesome. I hope that invigorates our audience to get into this field, so thank you for that as well. A little bit more personal, what is one thing about your job that keeps you coming back day after day, whether it's related to today's topic or not, what keeps you excited?

0:16:52.5 SC: I like learning new things and like with infectious disease, there's always exciting new developments happening, same with the AI. And so working in this interface, there's always something coming up every day. I love learning new things and with AI there are no barriers. Like I said, I always find something surprising or new that's coming up every day. So that's what I love.

0:17:17.6 SH: And lastly, what are some recent publications, projects, or initiatives that you're particularly excited about and would recommend that our listeners check out?

0:17:27.8 SC: Oh, sure. I would say probably the biggest poster child for AI in the biomedical space is AlphaFold, where I can predict protein structures from the sequence, which is like a big achievement. So, I would definitely ask people to check that out because it has big implications for drug discovery, for infections and other diseases. And I would say even for my lab's work, we've been trying to combine different types of data sets, like develop multimodal AI and mechanistic AI methods that we have been using for tuberculosis and other diseases. So that's another emerging area of AI, like trying to combine different data types that I want people to check out. Happy to provide links to you later.

0:18:13.9 SH: For sure. Yes, we'll share those resources with our audience down below the podcast episode. And I just wanna provide time if you would like to share anything else before we wrap up.

0:18:25.5 SC: No. Thank you so much for having me, really excited. If anyone wants to reach out with any questions or ideas, happy to discuss down the line.

0:18:35.2 SH: Dr. Chandrasekaran, thank you so much for joining us today to share from your experiences and provide some insight on this topic.

0:18:43.2 SH: To our listeners, we hope you've learned more about machine learning. We encourage you to check out the transcript and resources in the podcast notes. With that, we'll end here for today. Stay safe and stay prepared.