

AI-powered tool helps doctors detect rare diseases

July 3 2024, by Mary-Rose Abraham



Credit: Pixabay/CC0 Public Domain

In her first year at the David Geffen School of Medicine at UCLA, Katharina "Kat" Schmolly, MD, heard an old saying: "When you hear hoofbeats, think of horses, not zebras."

The medical maxim is a caution for physicians to prioritize likely causes rather than uncommon diagnoses. Dr. Schmolly, an undergraduate student of equine science and a former horse trainer, was on board.

But she began to reconsider during a hepatology lecture by Simon W. Beaven, MD, Ph.D. At his clinic in the Pfleger Liver Institute, Dr. Beaven treats patients with acute hepatic porphyria (AHP), a family of rare genetic diseases. Symptoms affect mostly women, often coinciding with the menstrual cycle, with severe, sometimes life-threatening, attacks that include abdominal pain, nausea and vomiting, limb weakness and anxiety.

"Women unfortunately get dismissed when they go to the emergency department over and over again for these complaints," said Dr. Schmolly. "Because it looks like it's menstrual pain, but actually, it could be a true liver disease."

She said the unfairness of it inspired her to found zebraMD, which uses artificial intelligence to help diagnose and manage rare and genetic diseases.

A predictive [algorithm](#) combs through electronic health records to identify disease patterns and flag patients who may be at risk so physicians can further test and diagnose. The model flips the medical maxim to highlight the rare zebras over commonly known horses.

"The diagnostic delay is roughly 10 to 15 years for these diseases because physicians don't see them very often," said Dr. Schmolly. "And while waiting for diagnosis, the disease can progress and cause irreversible damage."

"So, our goal is to diagnose patients earlier and manage their disease appropriately."

Not so rare

By definition, a rare disease affects less than 200,000 people. Rare diseases include more than 10,000 known conditions, and cumulatively affect more than 30 million people, or 1 in 10 Americans. That's about the same prevalence as diabetes.

A few, including multiple sclerosis, are well known. But the majority, such as bartonellosis, maple syrup urine disease and visual snow syndrome, are not.

AHP affects 1 in 100,000 people. The FDA approved givosiran in 2019 as a prophylactic treatment for recurrent attacks. But it takes an average of 15 years for diagnosis.

In an effort to speed that up, the drug's manufacturer, Alnylam Pharmaceuticals, approached researchers at the UCSF Porphyria Center about developing an algorithm to identify possible patients.

"Given that it's a rare disease, we knew up front that it was going to be difficult or impossible to run this as a single center endeavor. So we looked to our peers within the UC health network, and UCLA is the largest of the centers within UC health," said Vivek Rudrapatna, MD, Ph.D., an assistant professor in the division of gastroenterology at UCSF.

When Dr. Schmolly approached Dr. Beaven about her interest in rare diseases, he connected her with Dr. Rudrapatna. He is also director of The Real-World Evidence Lab, which applies data science techniques to electronic health records.

Together, they co-invented Project Zebra's predictive algorithm to analyze health care records and identify suspected porphyria patients. Dr. Rudrapatna had trained as a clinical data scientist during his UCSF

gastroenterology fellowship. Dr. Schmolly built on her pre-medical school experience with research and development at Medtronic where she worked on a diagnostic algorithm for an implantable cardiac monitoring device.

The data were de-identified patient records from UCSF and UCLA, a subset of the roughly 10 million records UC-wide. The first challenge for the algorithm: messy data.

Patient records comprise structured data, including vital signs, lab results and demographic information, and unstructured elements such as a physician typing up notes. Algorithms struggle with the latter. So, the researchers massaged the data, akin to organizing bunches of words into an orderly format.

"They're not going to learn much without some elbow grease," said Dr. Rudrapatna. "The Herculean challenge that was 90% of the effort for this study is going from unstructured clinical data to tightly organized, curated information that can be used to train statistical and machine learning algorithms."

Another challenge is masking certain [data points](#) that would allow the algorithm to "cheat" in making a disease prediction. Essentially, blinding the algorithm before a clinical suspicion of disease exists.

"At some point, a clinician refers a patient to [a porphyria specialist]," said Dr. Rudrapatna. "And then there's a referral order that pops up as structured data. Then the patient gets seen by the doctor. Then they get some blood test or urine test. Then they get confirmed for porphyria. The algorithm will see the referral and predict porphyria. If that's already happened, the algorithm is never going to learn to identify these patients earlier."

"What we want are algorithms that can really discover these patients way before clinicians are consciously thinking about it."

So what clues to AHP is the algorithm looking for?

Drs. Rudrapatna and Schmolly provided their model with three resources. The first was expert knowledge from Bruce Wang, MD, director of the UCSF Porphyria Center, who shared information on symptoms and symptom constellations. Second was providing access to a rare and genetic disease [database](#) from the National Institutes of Health which included signs, symptoms and presentation patterns for porphyria patients.

Third was allowing the algorithm to sift through the data and discover signals, allowing the data to speak for itself. This is especially important, Dr. Rudrapatna said, because rare diseases can be misdiagnosed so often.

"If you're a porphyria expert, you're thinking primarily about the signs and symptoms," he said. "But you might not be thinking about all the possible misdiagnoses these patients could have before they get diagnosed."

"An algorithm could find the tests that they're getting because people are moving down the wrong track. What are the erroneous therapies that they're receiving because they're misdiagnosed? Who are the providers they're seeing?"

In their [study](#) published in the *Journal of the American Medical Informatics Association*, the researchers found the algorithm predicted patients would be referred for AHP testing by a range of 89% to 93% accuracy. And when it came to predicting who tested positive for the disease, the algorithm recognized 71% of patients earlier than their actual diagnosis, corresponding to an average time saved of 1.2 years.

Privacy and permissions

Project Zebra is one of several AI platforms mining electronic health records. At UCLA Health, for example, researchers have [devised an algorithm](#) to identify patients with rare immune disorders. Another focuses on [early prediction](#) of end-stage kidney disease.

For any machine learning model, accessing patient data can often mean a lengthy approvals process. And since algorithms learn and perform better given massive datasets, health care records from a range of institutions would be ideal, though permissions would be a barrier.

To get around this, zebraMD uses Virtual Pooling, a patented technology developed by collaborator Trinabh Gupta, Ph.D., an assistant professor in the department of computer science at UC Santa Barbara. Virtual Pooling allows algorithms to learn from data without their explicit transfer.

"Owners of clinical data can maintain local control and security and privacy," said Dr. Rudrapatna, "but at the same time, allow machine learning models to learn from that collected experience, and then leave with those insights without leaving with the data itself."

An important part of training the AHP algorithm is validation by a specialist physician to ensure that its results are medically sound. Dr. Schmolly said the goal this year is to validate up to 350 diseases with an accuracy of at least 85%. That will provide a firm base to scale eventually to all 10,000 rare and [genetic diseases](#), even for extremely rare conditions which may not have a physician specialist.

Project Zebra is currently training its algorithm to predict cerebral aneurysms, a condition in which a weakened blood vessel in the brain balloons out or expands, possibly leading to its rupture and hemorrhage.

They are fairly common: about 1 in 50 to 1 in 100 Americans have an unruptured brain aneurysm. But rupture is a rare condition that affects about 30,000 people in the U.S. every year, with many of those fatalities. Diagnosing a brain aneurysm before it ruptures is key.

Geoffrey Colby, MD, Ph.D., is a professor of neurosurgery and radiology at the David Geffen School of Medicine at UCLA and the director of cerebrovascular neurosurgery at UCLA Health. He is advising the Project Zebra team as it develops a predictive algorithm for cerebral aneurysms, one of the diseases he devotes the most time to in his clinical practice.

He said if the algorithm flagged a patient, he would order imaging studies to look carefully at the blood vessels.

"The plus side is hopefully we can find people that have aneurysms and catch them before they have a problem," said Dr. Colby. "But we don't want the algorithm to identify a lot of people that would turn out to be false negatives. We don't want to incorrectly alarm lots of people."

He continued, "So my hope for this project is about identifying people who need help, and reducing the number of people who undergo a life-threatening event every year."

An eye on the future

Project Zebra will soon offer a web application. Working on that is Kristen Cardon, a software engineering intern and current doctoral candidate in the department of English at UCLA.

Cardon said she was inspired by her cousin who has a rare condition called Williams syndrome, characterized by cardiovascular disease and delays in cognitive development.

"Someone might come here looking for information about their own diagnosis," she said. "At a time that could be scary and alarming, our app is just making information accessible, easy and functional. I see it as a part of mental health justice as well as health justice more generally."

Physicians will use the app in Winter 2024 when zebraMD will test its algorithms in the real world by embedding into electronic health records systems at Ronald Reagan Medical Center at UCLA, Olive-View UCLA Medical Center, UCSF and Dartmouth Health.

"I hope that at some point this is a standard feature of any [electronic health records](#) system," said Dr. Schmolly, who will continue to lead the team during her internal medicine residency on a physician-scientist track at Dartmouth Health.

"The more patients we can diagnose, the more we can monitor over time. That means we can learn what works and what doesn't work. We can create precision medicine approaches for these rare diseases. Many of them don't have any treatments yet. So hopefully, we can find new treatment options for them."

More information: Balu Bhasuran et al, Reducing diagnostic delays in acute hepatic porphyria using health records data and machine learning, *Journal of the American Medical Informatics Association* (2024). [DOI: 10.1093/jamia/ocae141](https://doi.org/10.1093/jamia/ocae141)

Provided by University of California, Los Angeles

Citation: AI-powered tool helps doctors detect rare diseases (2024, July 3) retrieved 3 July 2024 from <https://medicalxpress.com/news/2024-07-ai-powered-tool-doctors-rare.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.