

Machine learning tool identifies rare, undiagnosed immune disorders through patients' electronic health records

May 1 2024



Credit: Pixabay/CC0 Public Domain

Researchers say a machine learning tool can identify many patients with

rare, undiagnosed diseases years earlier than current methods, potentially improving outcomes and reducing cost and morbidity. The findings, led by researchers at UCLA Health, are [described](#) in *Science Translational Medicine*.

"Patients who have rare diseases may face prolonged delays in diagnosis and treatment, resulting in unnecessary testing, progressive illness, psychological stresses, and financial burdens," said Manish Butte, MD, Ph.D., a UCLA professor in pediatrics, human genetics, and microbiology/immunology who cares for these patients in his clinic at UCLA.

"Machine learning and other artificial intelligence methods are making their way into health care. Using these tools, we developed an approach to speed the diagnosis of undiagnosed patients by identifying patterns in their [electronic health records](#) that resemble those of patients who are known to have the disorders."

This study focused on disorders collectively called common variable immunodeficiency (CVID), which often elude diagnosis for years or decades after symptom onset because the disorders are rare, symptoms can vary greatly from person to person, and symptoms tend to overlap with those of other more common disorders.

Additionally, the disorders in each individual are often driven by changes in only one gene—but not the same gene from one manifestation of the disorder to another—and over 60 genes have been implicated thus far. Without a single causal mechanism, there are no genetic tests to provide a definitive diagnosis.

CVID is one of the most common human inborn errors of immunity (IEI)—rare diseases that increase a person's susceptibility to infection, autoimmunity and autoinflammation. More than 500 IEIs have been

identified, and more are discovered each year. CVID, estimated to affect 1 in 25,000 people, is associated with antibody deficiencies—both quantity and function—and impaired immune responses.

Butte and Bogdan Pasaniuc, Ph.D., a professor of computational medicine, [human genetics](#), and pathology and laboratory medicine at UCLA David Geffen School of Medicine, led a team that developed a [machine learning](#) tool called PheNet, borrowing from the term "phenotypes," the observable characteristics or traits of a disease as seen in an individual. PheNet learns phenotypic patterns from verified CVID cases and uses this knowledge to rank patients by likelihood of having CVID.

"The clinical presentation of rare immune phenotypes such as CVID intersects with many medical specialties. Patients may be seen in ear, nose and throat clinics for sinus infections. They may be treated in pulmonology clinics for pneumonias. This fragmentation of care across multiple specialists leads to long delays in diagnosis and treatment," said Butte, who with Pasaniuc, is a co-senior author of the journal article.

It's impossible to teach all these busy specialists about immune deficiencies with the hopes that even if they could recognize which patients have an underlying immune defect, they'll refer these patients to us. We had to find a better way to find these patients."

"Our own patients report experiencing years to decades of symptoms before they were referred to our immunology clinic," Butte added.

"With PheNet, dozens of patients could have been diagnosed one to four years earlier than they were, and by bringing patients to care years earlier, we should be able to reduce their costs and improve their health outcomes."

Because there is no single clinical presentation for CVID, identifying an

electronic health record "signature" for the disorder is not straightforward. The researchers developed their computational algorithm to infer EHR signatures from the records of patients known to have CVID and from the patterns of illnesses found in the literature.

The software then computes a numerical score for each patient that rank orders the patients most likely to have CVID. Those with high scores—patients who the researchers describe as "hiding in the [medical system](#)"—would be candidates for referral to an immunology specialist.

Pasaniuc said that when the research team applied PheNet to the UCLA electronic health record data comprising millions of patient records and followed up with a blinded chart review of the top 100 patients ranked by the system, they found that 74% were deemed probable to have CVID. Based on these preliminary data, Butte and Pasaniuc have begun to apply their AI in the real world.

They started by validating PheNet with more than 6 million records of patients from disparate medical systems in the University of California Data Warehouse and at Vanderbilt Medical Center in Tennessee. A collaboration led by Butte to have specialists see the patients identified by the algorithm was launched with the immunology clinics at University of California campuses in San Diego, Irvine, Davis, and San Francisco.

"We show that artificial intelligence algorithms such as PheNet can offer clinical benefits by expediting the diagnosis of CVID, and we expect this to apply to other [rare diseases](#), as well," Pasaniuc said.

"Our implementation across all five University of California medical centers is already making an impact. We are now improving the precision of our approach to better identify CVID while expanding to other diseases. We will also plan to teach the system to read medical notes to glean even more information about patients and their illnesses."

Lead author Ruth Johnson, Ph.D., a former member of the Pasaniuc Lab who now is a fellow at Harvard Medical School, said limitations of the current health care system can result in tunnel vision, where different doctors see different aspects of a disease but are unable to put the whole picture together. This delays diagnosis, especially for the many CVID patients who have multisystem manifestations that fluctuate over time. Artificial intelligence can overcome these obstacles.

"For every year a diagnosis is delayed, there is an increase in infections, antibiotic use, emergency room visits, hospitalizations, and missed days of work and school," she said. "In addition to the financial and emotional toll this takes on patients and their families, the aggregate impact to the U.S. health system of failing to diagnose CVID in a timely fashion is likely to be in the millions or billions of dollars."

In addition to Butte and Pasaniuc, UCLA authors include Ruth Johnson (first author), Alexis V. Stephens, Rachel Mester, Sergey Knyazev, Lisa A. Kohn, Malika K. Freund, Leroy Bondhus, Brian L. Hill, Tommer Schwarz, Noah Zaitlen, and Valerie A. Arboleda. Lisa A. Bastarache contributed from the Department of Biomedical Informatics at Vanderbilt University.

More information: Ruth Johnson et al, Electronic health record signatures identify undiagnosed patients with common variable immunodeficiency disease, *Science Translational Medicine* (2024). [DOI: 10.1126/scitranslmed.ade4510](https://doi.org/10.1126/scitranslmed.ade4510)

Provided by University of California, Los Angeles

Citation: Machine learning tool identifies rare, undiagnosed immune disorders through patients' electronic health records (2024, May 1) retrieved 12 September 2024 from

<https://medicalxpress.com/news/2024-05-machine-tool-rare-undiagnosed-immune.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.