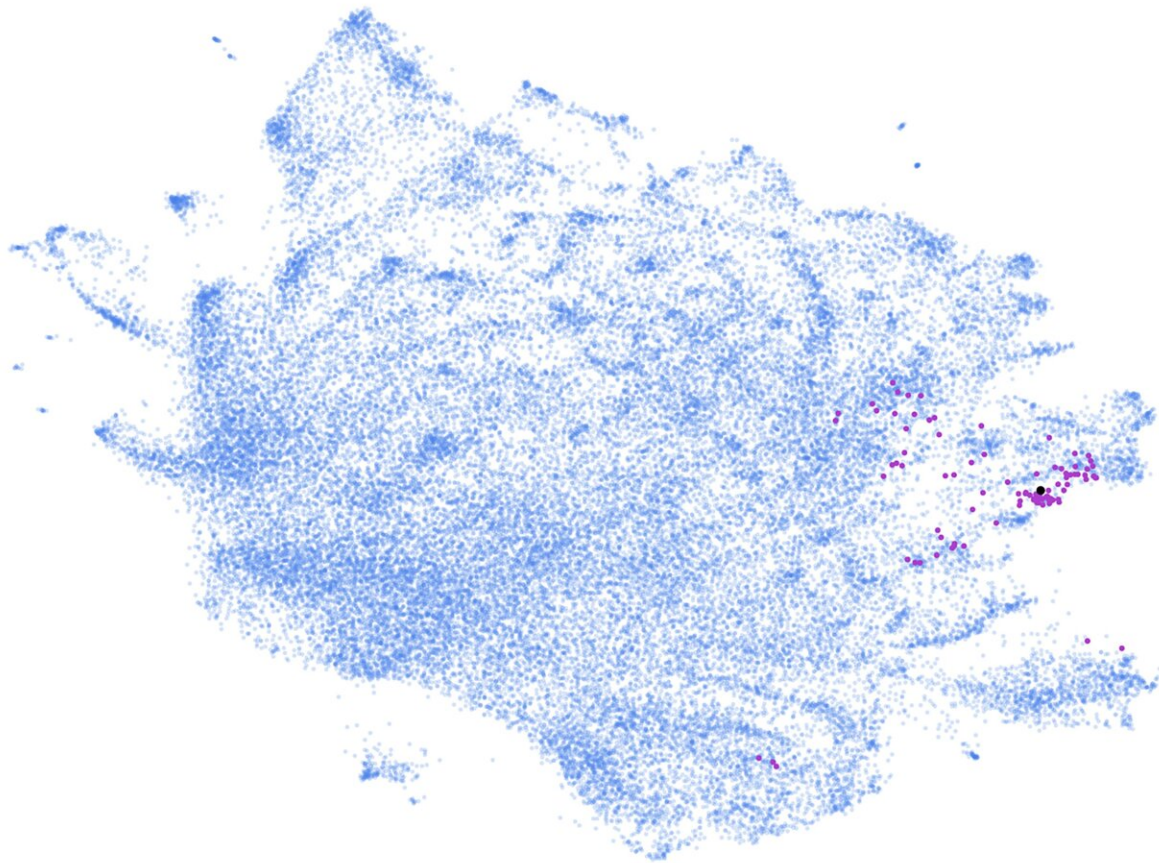


Scientists create a labor-saving automated method for studying electronic health records

September 2 2021



Mount Sinai scientists created an AI-based, automated system that learns to read patient data from electronic health records. Here the system identified dementia cases (purple dots) from a database of nearly 2 million patients (blue dots).

Credit: Glicksberg lab, Mount Sinai, N.Y., N.Y.

In an article published in the journal *Patterns*, scientists at the Icahn School of Medicine at Mount Sinai described the creation of a new, automated, artificial intelligence-based algorithm that can learn to read patient data from electronic health records. In a side-by-side comparison, they showed that their method, called Phe2vec (FEE-to-vek), accurately identified patients with certain diseases as well as the traditional, "gold-standard" method, which requires much more manual labor to develop and perform.

"There continues to be an explosion in the amount and types of data electronically stored in a patient's medical record. Disentangling this complex web of data can be highly burdensome, thus slowing advancements in [clinical research](#)," said Benjamin S. Glicksberg, Ph.D., Assistant Professor of Genetics and Genomic Sciences, a member of the Hasso Plattner Institute for Digital Health at Mount Sinai (HPIMS), and a senior author of the study. "In this study, we created a new method for mining data from electronic health records with machine learning that is faster and less labor intensive than the industry standard. We hope that this will be a [valuable tool](#) that will facilitate further, and less biased, research in clinical informatics."

The study was led by Jessica K. De Freitas, a graduate student in Dr. Glicksberg lab.

Currently, scientists rely on a set of established [computer programs](#), or algorithms, to mine medical records for new information. The development and storage of these algorithms is managed by a system called the Phenotype Knowledgebase (PheKB). Although the system is highly effective at correctly identifying a patient diagnosis, the process of developing an algorithm can be very time-consuming and inflexible. To study a disease, researchers first have to comb through reams of medical records looking for pieces of data, such as certain lab tests or prescriptions, which are uniquely associated with the disease. They then

program the algorithm that guides the computer to search for patients who have those disease-specific pieces of data, which constitute a "phenotype". In turn, the list of patients identified by the computer needs to be manually double-checked by researchers. Each time researchers want to study a new disease, they have to restart the process from scratch.

In this study, the researchers tried a different approach—one in which the computer learns, on its own, how to spot disease phenotypes and thus save researchers time and effort. This new, Phe2vec method was based on studies the team had already conducted.

"Previously, we showed that unsupervised machine learning could be a highly efficient and [effective strategy](#) for mining electronic health records," said Riccardo Miotto, Ph.D., a former Assistant Professor at the HPIMS and a senior author of the study. "The potential advantage of our approach is that it learns representations of diseases from the data itself. Therefore, the machine does much of the work experts would normally do to define the combination of data elements from health records that best describes a particular disease."

Essentially, a computer was programmed to scour through millions of electronic health records and learn how to find connections between data and diseases. This programming relied on "embedding" algorithms that had been previously developed by other researchers, such as linguists, to study word networks in various languages. One of the algorithms, called word2vec, was particularly effective. Then, the computer was programmed to use what it learned to identify the diagnoses of nearly 2 million patients whose data was stored in the Mount Sinai Health System.

Finally, the researchers compared the effectiveness between the new and the old systems. For nine out of ten diseases tested, they found that the

new Phe2vec system was as effective as, or performed slightly better than, the gold standard phenotyping process at correctly identifying a diagnoses from [electronic health records](#). A few examples of the diseases included dementia, multiple sclerosis, and [sickle cell anemia](#).

"Overall our results are encouraging and suggest that Phe2vec is a promising technique for large-scale phenotyping of diseases in electronic health record data," Dr. Glicksberg said. "With further testing and refinement, we hope that it could be used to automate many of the initial steps of clinical informatics research, thus allowing scientists to focus their efforts on downstream analyses like predictive modeling."

More information: De Freitas, J.K., et al., Phe2vec: Automated Disease Phenotyping based on Unsupervised Embeddings from Electronic Health Records, *Patterns*, September 2, 2021, [DOI: 10.1016/j.patter.2021.100337](#)

Provided by The Mount Sinai Hospital

Citation: Scientists create a labor-saving automated method for studying electronic health records (2021, September 2) retrieved 5 May 2024 from <https://medicalxpress.com/news/2021-09-scientists-labor-saving-automated-method-electronic.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.