



**Mount
Sinai**

DRAFT

Media Contact: Christopher Thomas

Mount Sinai Press Office

347-541-0212

NewsMedia@mssm.edu

Embargoed for release:

September 2, 2021

11:00 am EDT

Scientists Create a Labor-Saving Automated Method for Studying Electronic Health Records

Mount Sinai study suggests new method is as effective as manually-based “gold-standard” at classifying a diagnosis

New York, NY (September 2, 2021) – 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, PhD, 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



Reading Dementia

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). *Courtesy of the Glicksberg lab, Mount Sinai, N.Y., N.Y.*

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, PhD, 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.”

This study was supported by the Hasso Plattner Foundation, the Alzheimer’s Drug Discovery Foundation, and a courtesy graphics processing unit donation from the NVIDIA Corporation.

Article

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.

About the Mount Sinai Health System

The Mount Sinai Health System is New York City's largest academic medical system, encompassing eight hospitals, a leading medical school, and a vast network of ambulatory practices throughout the greater New York region. Mount Sinai advances medicine and health through unrivaled education and translational research and discovery to deliver care that is the safest, highest-quality, most accessible and equitable, and the best value of any health system in the nation. The Health System includes approximately 7,300 primary and specialty care physicians; 13 joint-venture ambulatory surgery centers; more than 415 ambulatory practices throughout the five boroughs of New York City, Westchester, Long Island, and Florida; and more than 30 affiliated community health centers. The Mount Sinai Hospital is ranked on *U.S. News & World Report's* "Honor Roll" of the top 20 U.S. hospitals and is top in the nation by specialty: No. 1 in Geriatrics and top 20 in Cardiology/Heart Surgery, Diabetes/Endocrinology, Gastroenterology/GI Surgery, Neurology/Neurosurgery, Orthopedics, Pulmonology/Lung Surgery, Rehabilitation, and Urology. New York Eye and Ear Infirmary of Mount Sinai is ranked No. 12 in Ophthalmology. Mount Sinai Kravis Children's Hospital is ranked in *U.S. News & World Report's* "Best Children's Hospitals" among the country's best in four out of 10 pediatric specialties. The Icahn School of Medicine is one of three medical schools that have earned distinction by multiple indicators: ranked in the top 20 by *U.S. News & World Report's* "Best Medical Schools," aligned with a *U.S. News & World Report* "Honor Roll" Hospital, and No. 14 in the nation for National Institutes of Health funding. *Newsweek's* "The World's Best Smart Hospitals" ranks The Mount Sinai Hospital as No. 1 in New York and in the top five globally, and Mount Sinai Morningside in the top 20 globally.

For more information, visit <https://www.mountsinai.org> or find Mount Sinai on [Facebook](#), [Twitter](#) and [YouTube](#).

###