A machine learning tool developed at Carnegie Mellon University uses big data from the electronic health record to accurately predict sepsis.

Researchers at Carnegie Mellon University’s (CMU) Heinz College are applying a machine learning algorithm to big data in the electronic health record (EHR) to more accurately predict sepsis, one of the most dangerous and insidious hospital acquired conditions.

The Sepsis Alliance reports that more than 1.7 million people in the US are diagnosed with sepsis annually. Of those affected by the condition, an estimated 270,000 die each year.

Sepsis is also the number one driver of hospital costs in the US, consuming more than $27 billion annually. Many times, the infection is acquired in the community, and patients with complex comorbidities are often at the highest risk.

“The problem underlying sepsis is that it’s incredibly heterogeneous,” said Jeremy Weiss, MD, PhD, Assistant Professor of Health Informatics at CMU’s Heinz College, to HealthITAnalytics.com

“Anybody can get sepsis from a multitude of infections, at different sites, and with different comorbid profiles.”

Traditionally, providers identify high-risk sepsis patients by evaluating symptoms and medical histories.

However, Weiss and his team are working to improve the speed and accuracy of sepsis prediction.

Using machine learning and EHR data, Weiss has developed a method of accurately assigning risk scores to patients, offering a way to catch sepsis earlier than is possible with standard processes.

“EHR data is very detailed. There’s a lot of time-stamped information,” Weiss said.

“A lot of classical analyses don’t get to capture that kind of information. With EHRs, where this data is automatically entered, we can look at the temporal progression of disease and update our risk models more adeptly.”

Weiss and his team utilize this time-stamped data to evaluate information such as blood tests, prescribed drugs, and blood pressure. This data is typically contained in the structured portion of the EHR, which they can access during a healthcare encounter and use to make real-time predictions.

“Data such as lab tests, prescribed medications, and vital signs will inform us when procedures were performed and the background set of diagnoses for the patients upon entry,” said Weiss.

To extract relevant patterns from structured EHR information, Weiss and his team are applying machine learning algorithms, which can analyze many different data points and assist in making accurate predictions, he said.

Weiss and his team are also exploring how clustering of similar cases can help to refine predictive analytics and get ahead of sepsis that may develop in future patients.

“If we can identify clusters with very specific phenotypes, then we can tailor treatments to those subgroups,” said Weiss.

The Heinz College group is not the first to use machine learning algorithms and EHR data to accurately predict sepsis in patients.

Researchers at the University of Pennsylvania developed a machine learning tool that continuously monitored EHRs and identified patients headed for sepsis or septic shock a full 12 hours before the onset of the condition.

Additionally, researchers at North Carolina State University, in collaboration with Mayo Clinic and Christiana Care Health System in Delaware, have used EHR data and machine learning to improve how the healthcare system identifies and treats patients with sepsis.

Weiss and his team hope to make similar strides in sepsis prediction and prevention.

“We’re trying to leverage much more information from the EHR to tailor simpler risk scores,” Weiss said.

“While those scores are a really good, quick check for what your general risk is, we’re really interested in having a much better risk estimate that will draw from a lot more signals.”