CASE STUDY

Health and Life Sciences
Data Driven

Using Machine Learning and EMR Data to Predict Patient Decline

Predictive clinical analytics with technologies from Intel and Cloudera* help Sharp HealthCare* use electronic medical record data to identify patients at risk of a sudden decline

At a Glance

Sharp HealthCare* used technologies from Intel and Cloudera* to explore predictive clinical analytics and machine learning. The new methods show potential to help:

- Empower rapid-response teams to intervene proactively
- Improve the quality and cost of care
- Enhance resource utilization
- Deliver more value from investments in electronic medical records

"With surprising accuracy, we found you really can predict when a patient is heading in the wrong direction, just from analyzing the EMR data that is available in real time . . . There is a lot of potential for this type of technology to deliver a clinical and financial return on investment."

—Brett MacLaren
Vice President of Enterprise Analytics
Sharp HealthCare

Can predictive analytics alert a hospital's rapid-response team (RRT) to a health crisis in the making? Yes they can, said Brett MacLaren, vice president of enterprise analytics at Sharp HealthCare in San Diego, California. MacLaren oversaw an eight-week proof-of-concept (PoC) project that analyzed data from the hospital's Cerner* electronic medical record (EMR) system to identify patients who were at risk of requiring an intervention from the rapid response team within the next hour. He said the PoC highlights the potential for predictive analytics to help hospitals improve the quality and cost of patient care and make more effective use of highly skilled clinicians.

Challenge

Sharp HealthCare has established specialized rapid response teams that respond to medical emergencies in the hospital. As time permits, these teams may manually review charts to spot a potential crisis. Sharp wanted to see if predictive analytics techniques, including artificial intelligence methods such as machine learning, could automate that analysis and help the hospital deliver more proactive, efficient interventions.

Solution

Sharp worked with Intel and ProKarma, Inc.* on a PoC that used data from Sharp's Cerner EMR system to create a predictive model intended to detect patient health decline and identify patients at risk for requiring a rapid response team intervention within a defined window of time. The team selected multiple features, including blood pressure, temperature, and pulse rate, and used machine learning to train the model. The scalable solution ran on a Cloudera* cluster powered by the Intel® Xeon® processor E5 v4 family.

Results

When Sharp tested the model against historical data, the model was 80-percent accurate[1] in predicting the likelihood of an RRT event within an hour, demonstrating the potential to drive real-time clinical interventions, improve outcomes, and enhance resource utilization. With distributed analytics and scale-out Intel® architecture, Sharp has identified an actionable way to capture additional value from its previous investments in EMR technology.

Leveraging Existing Data to Help Improve Care

Hospitals have made deep investments in their EMR systems. Now, leaders like Sharp HealthCare are finding new ways to extract more value from those systems.

"Healthcare is starting to get beyond traditional transactional decision making, and move toward real-time, predictive, interventional decision making at the point of care," said Sharp's Brett MacLaren. "We're beginning to use analytics not just to understand what happened in the past and make operational decisions, but to predict the future and intervene in real time to influence the clinical outcome."

MacLaren views Sharp's PoC as a key step to moving the hospital system up the analytics maturity curve. "Because of the importance of predictive clinical analytics, we want to build our in-house competency," he said. "We decided to start with a proof of

concept to explore whether the patient data in the current EMRs could accurately predict if any patients would be likely to need an RRT intervention. We had RRT people who were doing some manual chart review for the sickest patients, but we wanted to see if we could apply some data science and give them a tool to predict an RRT event so they could intervene more proactively."

A Direct Path to Predictive Analytics

Using EMRs gave Sharp a direct and affordable starting point for predictive analytics. In addition to capitalizing on existing data, this approach allowed the hospital to use scalable Intel architecture-based infrastructure and open source tools such as Hadoop* and Apache Spark*.

To build the model, the PoC team loaded structured data from the clinical data warehouse into a nine-node Cloudera EDH Hadoop cluster based on the Intel Xeon processor E5 v4 family. The team used Apache Impala* to query Cerner data in the Apache Hive* format, and conducted the analysis and modeling using Anaconda* in Jupyter Notebooks*—all part of the Cloudera Enterprise suite. Team members developed and refined their predictive scoring model throughout the eight-week PoC, using supervised machine learning to train the model and comparing real-time data recorded in the EMRs against a sandbox data set.

The model performed with 80-percent overall accuracy and 82-percent precision,[2] indicating the presence of predictive signals in the data. "We were pleasantly surprised at the predictability and accuracy of the model even with a relatively limited data set," MacLaren said.

Refining the Model

The PoC yielded a number of ways to refine the model, including providing a longer period of historical data for model training and incorporating additional health data such as lab results and medication interactions.

There were also surprises. Demographic data such as age and gender, which many historical models and patient risk systems discard, turned out to be an important predictive factor. Future model development could also explore the use of other non-EMR data streams, such as socioeconomic and geospatial data. "We know these factors influence remissions, and they may well be relevant in driving clinical health also," MacLaren said.

In addition, data from the Internet of Things will dramatically increase the accuracy of the model, according to MacLaren. "Things like manually entered vital signs are not recorded with sufficient consistency to effectively predict an RRT event," he said. "By leveraging actual, real-time machine data and log files, we'll get denser and more consistent data and a vastly better ability to predict."

Spotlight on Sharp HealthCare

Sharp HealthCare is San Diego's leading healthcare provider, serving more people in the San Diego area than any other healthcare system. Sharp has earned its leadership position through its commitment to delivering the extraordinary level of care that characterizes "the Sharp experience." The not-for-profit system has more than 2,900 affiliated physicians and 18,000 employees, working at four acute-care hospitals, three specialty hospitals, three affiliated medical groups, and a full spectrum of other facilities and services. Sharp is accredited by the Joint Commission, and had net revenues in 2015 of USD 3.4 billion. Learn more at **www.sharp.com**.

Replacing Hunches with Data-Enabled Decisions

Sharp's leaders see broad value from predictive clinical analytics, whether to identify patients at risk for developing life-threatening sepsis, a longer-than-average length of stay, or a costly readmission.

"There is a lot of potential for this type of technology to deliver a clinical and financial return on investment," MacLaren said. "At Sharp, we never want to miss an opportunity to intervene to improve care. If we can give the RRT team a one-hour head start and they can prevent a near-mortality event from occurring, we can reduce mortality. We can avoid all the negative consequences of a near-mortality event, and make possible a whole host of benefits to patient care and outcomes. We also improve efficiency and make better use of the rapid response team. We'd be analyzing the available data automatically and giving them actionable patient risk scores instead of forcing them to comb the chart and act on hunches."

Modernizing the analytics infrastructure is part of Sharp's preparation for large-scale health analytics. "We need to move beyond just having an effective yet limited relational database management infrastructure that was developed in the 1990s, and migrate to a twenty-first-century data model to more effectively support data-driven decision making," he said. Sharp's three-year modernization roadmap includes an uplift of the data warehouse to create a data lake that will provide wider access to diverse types of data.

"Analytics tools are improving at an unbelievably fast pace," MacLaren said. "I believe if you have a good understanding of the problem you're trying to solve, the tools will be there to develop and run the solution. Part of our job now is to evaluate and redesign our infrastructure so we can be ready to use these tools—to provide more proactive care and create a data-driven healthcare enterprise."

Find the solution that is right for your organization. Contact your Intel representative or visit **intel.com/healthcare**.

[1] 80-percent accuracy indicates the level of accuracy observed when scoring a set of unlabeled test data that was not used in the development of the model.

[2] 62-percent precision indicates the level of precision observed when scoring a set of unlabeled test data that was not used in the development of the model.