Google Cloud

quantiphi

Using AI and Machine Learning to automate cancer recurrence tagging

Cancer recurrence tagging using patient's historical EHR records, manual abstraction practices, and legacy data subsystems can take upto a year. A cloud-based solution from Quantiphi can do the same thing in minutes.

About Quantiphi

Based in Marlborough, MA, Quantiphi is an award-winning Applied AI and data science software and services company driven by the desire to solve transformational problems at the heart of business. It is Google Cloud's Global Social Impact Partner of the Year for 2019 and Machine Learning Partner of the Year for 2017 and 2018.

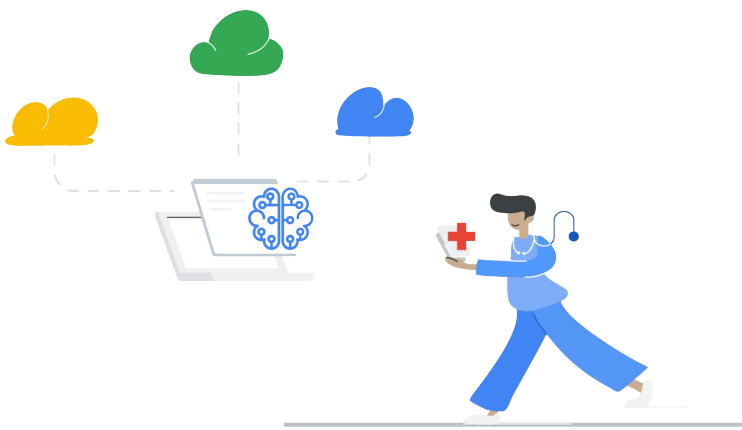
Quantiphi is passionate about its customers and obsessed with problem-solving to make products smarter, customer experiences frictionless, processes autonomous and businesses safer by detecting risks, threats and anomalies. Healthcare is one of its strongest and fastest-growing business segments.

Identifying patients where cancer has recurred is like finding needles in a field of haystacks. For a large, multi-state non-profit health system in need of a solution, cancer recurrence represented less than one percent of oncology related encounters.

“One of the major challenges in the field of cancer research today is the availability of real world clinical evidence for researchers,” says Asif Hasan, co-founder of Quantiphi, a Google Cloud partner specializing in Artificial Intelligence, Machine Learning and all things Data. Researchers building machine learning models in oncology need to know, among other things, each patient’s incoming diagnosis, when their cancer was first detected, what treatment protocols followed, what medications they took, and what interventions doctors tried, as well as the eventual outcome.

“It’s hugely important for researchers to have this entire information in cancer registries,” Hasan continues. The more quality data they have, the better their chances of tagging—based on insight collected from other cases—that a patient’s cancer is likely to come back. Quantiphi quickly learned that gathering all the necessary data and curating features from a mix of structured and unstructured data would be a massive undertaking for its client.

Using Natural Language processing techniques, more than 1,300 data features of patient encounters were generated for modeling cancer recurrence that cancer researchers may find useful. Multiplying this number by millions of patient encounters would result in billions of data points.



Google Cloud

quantiphi

Since not every data point applies to every patient, the data matrix contained a lot of missing values. Adding to the challenge, records that covered patient encounters, including pathology reports, medications, diagnoses, treatment plans, physician notes in free-form text across multiple Electronic Health Record (EHR) systems were not necessarily consistent from one visit to the next, much less from one patient to the next. Complicating matters even further, the data model must also account for the sequence of a patient's visits and how often they visited, which could be five times in one month or five times in one year.

Using machine learning at scale using cloud to automate tagging of cancer recurrence

Before researchers could apply analytics to help tag cancer recurrence, patient data across the provider's oncology population had to be captured in a consistent format, a process Hasan describes as "hugely tedious and very time consuming."

"Think of a clinical expert," he suggests, "maybe a trained nurse or a physician looking at reams and reams of information in electronic health records, looking at labs, medications, and freeform texts that various clinicians have entered, and then from there, trying to abstract information that eventually needs to make its way into the cancer registry - It's a huge burden on the healthcare system. Every minute the clinician is out there abstracting data from electronic health records and putting it into a cancer registry is a minute they could have spent delivering patient care."

Overall, hundreds of thousands of oncology patients had millions of visits. "There's a lot of heterogeneity in the interactions that we have had," says Hasan. "Our job is to then take all of this information, and from there, automate the process of detecting the recurrence of cancer from the electronic health record data that's made available to us."

Asif Hasan,
Co-founder, Quantiphi

"We can leverage the power of machine learning on Google Cloud to solve a very sizable challenge in cancer research, making real-world evidence available for researchers without having experts spend valuable time abstracting this data manually and entering it into the system."

A large part of data ingestion was undertaken by the client's data engineering team. Quantiphi recommends leveraging Google Cloud Healthcare API for this task in the future. "Cloud Healthcare API bridges the gap between care systems and Google Cloud," explains Shantam Gupta, machine learning engineer at Quantiphi. "Using the Cloud Healthcare API, you can ingest structured and semi-structured data. You can create, store, import, export, and manage FHIR, HL7v2 and DICOM stores. You can also de-identify your data to protect personally identifiable information. It also helps in data persistence. Overall, this API makes it easier and faster to get actionable healthcare insights in a compliance-heavy environment."

Google Cloud

quantiphi

The Cloud Healthcare API connects EHR data sources and Google Cloud. Once the data has been ingested into Google Cloud Platform, managed services like BigQuery and Dataproc can greatly reduce computational workload and simplify complex operations to prepare big data for downstream modeling and analytics. De-identified longitudinal patient record prepared for analysis stays in Google Cloud, where it is consumed by ML models on AI Platform for tagging cancer recurrence. Information on a patient's cancer recurrence state is now available to the entire health system, making it easier for different departments to optimize care coordination.

Saving time for the experts while augmenting their expertise

The system Quantiphi has developed in partnership with Google Cloud has reduced the process of tagging cancer recurrence from as much as a year to a matter of minutes. "We can leverage the power of machine learning on Google Cloud to solve a very sizable challenge in cancer research, making real-world evidence available for researchers without having experts spend valuable time abstracting this data manually and entering it into the system," says Hasan.

Shantam Gupta, Machine Learning Engineer, Quantiphi

"Overall, the Google Cloud Healthcare API will make it easier and faster to get actionable healthcare insights with its HIPAA-compliant services and environment."

Not only has their work produced important cost savings by redirecting valuable clinician expertise away from data mining to patient care, but it has also demonstrated the power of augmenting clinical wisdom with machine learning. "Doing this project makes us very, very optimistic," says Hasan, "that when we combine the intuition of clinicians with assistive AI techniques and use the power of Google Cloud to solve some of these computational challenges, we can really impact the healthcare system in a very positive way."

To learn more about Google Cloud Healthcare and Life Sciences visit:

https://cloud.google.com/solutions/healthcare-life-sciences

To learn more about Quanitphi visit:

https://www.quantiphi.com