# New AI technology integrates multiple data types to predict cancer outcomes
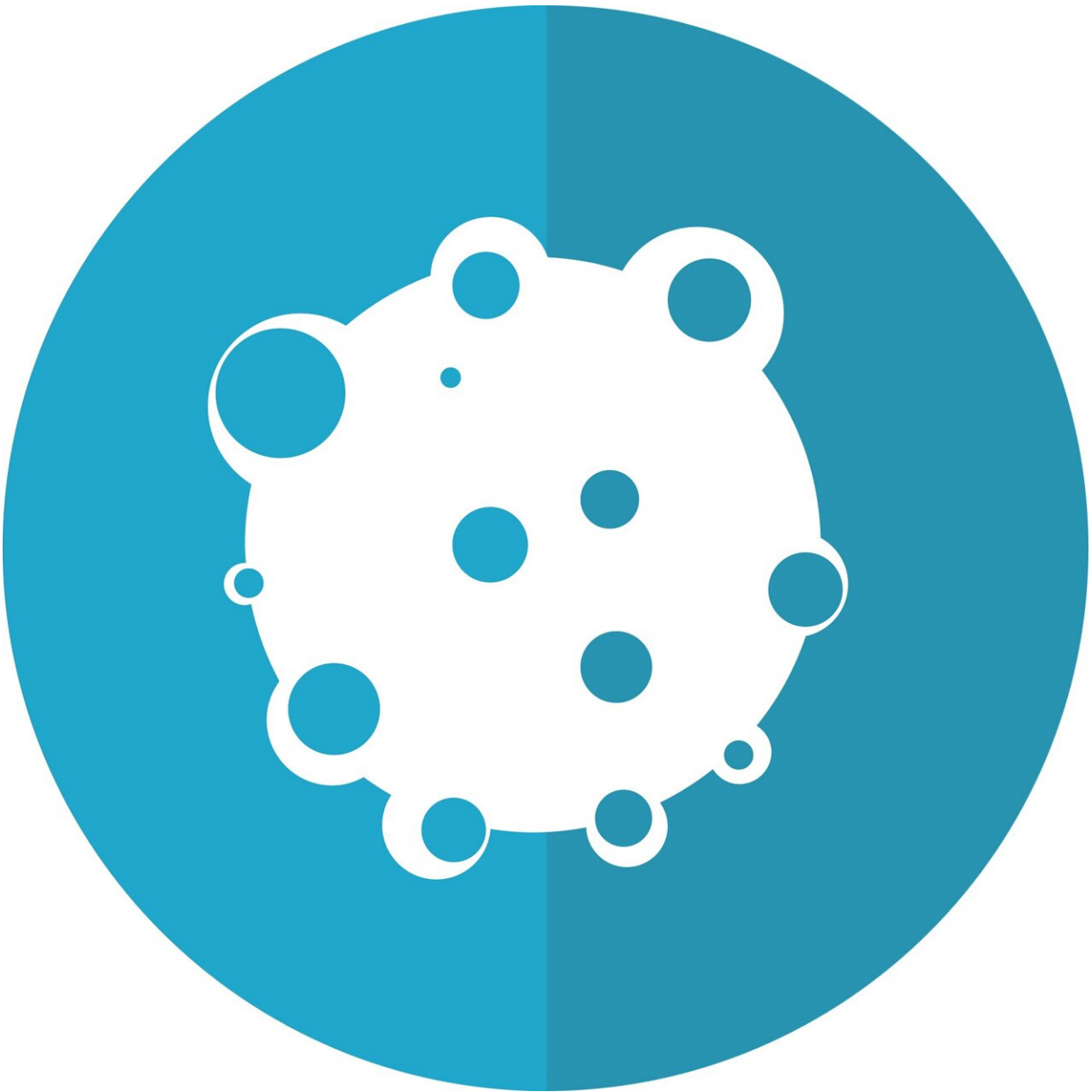
August 8 2022

While it's long been understood that predicting outcomes in patients with cancer requires considering many factors, such as patient history, genes and disease pathology, clinicians struggle with integrating this information to make decisions about patient care. A new study from researchers from the Mahmood Lab at Brigham and Women's Hospital reveals a proof-of-concept model that uses artificial intelligence (AI) to combine multiple types of data from different sources to predict patient outcomes for 14 different types of cancer. Results are published in *Cancer Cell*.

Experts depend on several sources of data, like genomic sequencing, pathology, and patient history, to diagnose and prognosticate different types of cancer. While existing technology enables them to use this information to predict outcomes, manually integrating data from different sources is challenging and experts often find themselves making subjective assessments.

"Experts analyze many pieces of evidence to predict how well a patient may do," said Faisal Mahmood, Ph.D., an assistant professor in the Division of Computational Pathology at the Brigham and associate member of the Cancer Program at the Broad Institute of Harvard and MIT. "These early examinations become the basis of making decisions about enrolling in a clinical trial or specific treatment regimens. But that means that this multimodal prediction happens at the level of the expert. We're trying to address the problem computationally."

Through these new AI models, Mahmood and colleagues uncovered a means to integrate several forms of diagnostic information computationally to yield more accurate outcome predictions. The AI

models demonstrate the ability to make prognostic determinations while also uncovering the predictive bases of features used to predict patient risk—a property that could be used to uncover new biomarkers.

Researchers built the models using The Cancer Genome Atlas (TCGA), a publicly available resource containing data on many different types of cancer. They then developed a multimodal deep learning-based algorithm which is capable of learning prognostic information from multiple data sources. By first creating separate models for histology and genomic data, they could fuse the technology into one integrated entity that provides key prognostic information. Finally, they evaluated the model's efficacy by feeding it data sets from 14 cancer types as well as patient histology and genomic data. Results demonstrated that the models yielded more accurate patient outcome predictions than those incorporating only single sources of information.

This study highlights that using AI to integrate different types of clinically informed data to predict disease outcomes is feasible. Mahmood explained that these models could allow researchers to discover biomarkers that incorporate different clinical factors and better understand what type of information they need to diagnose different types of cancer. The researchers also quantitively studied the importance of each diagnostic modality for individual cancer types and the benefit of integrating multiple modalities.

The AI models are also capable of elucidating pathologic and genomic features that drive prognostic predictions. The team found that the models used patient immune responses as a prognostic marker without being trained to do so, a notable finding given that previous research shows that patients whose tumors elicit stronger immune responses tend to experience better outcomes.

While this proof-of-concept model reveals a newfound role for AI

technology in cancer care, this research is only a first step in implementing these models clinically. Applying these models in the clinic requires incorporating larger data sets and validating on large independent test cohorts. Going forward, Mahmood aims to integrate even more types of patient information, such as radiology scans, family histories, and electronic medical records, and eventually bring the model to clinical trials.

"This work sets the stage for larger health care AI studies that combine data from multiple sources," said Mahmood. "In a broader sense, our findings emphasize a need for building computational pathology prognostic models with much larger datasets and downstream clinical trials to establish utility."

  **More information:** Faisal Mahmood, Pan-Cancer Integrative Histology-Genomic Analysis via Multimodal Deep Learning, *Cancer Cell* (2022). DOI: 10.1016/j.ccell.2022.07.004. www.cell.com/cancer-cell/fullt … 1535-6108(22)00317-8

Provided by Brigham and Women's Hospital