Cancerworld

Machine learning improves accuracy of survival prediction

Janet Fricker / 2 November 2023

An artificial intelligence prototype has been developed to estimate the survival of newly diagnosed cancer patients with breast, thyroid, and pancreatic cancers. The study, abstract 4160, presented at the American College of Surgeons (ACS) Clinical Congress 2023, October 22–25, Boston, found that age at diagnosis, tumour size, whether the patient had surgery, and time from diagnosis to treatment, were all factors that needed to be taken into consideration in the calculation.

"We sought to develop this Cancer Survival Calculator to provide a more personalised estimate of what patients can expect regarding their cancer prognosis," says lead author, Lauren Janczewski, from Northwestern University McGaw Medical Center, Chicago.

Traditionally, cancer survival is estimated according to tumour stage; however, patient survival is known to be dependent on multiple factors including individual patient characteristics, treatment used, and additional tumour specific factors. "Having the ability to accurately prognosticate for cancer patients is critical to providing high quality care in order to counsel them on what they can expect regarding their cancer diagnosis, increase patient education, and engage in shared decision making," Janczewski tells *Cancerworld*. The team sought to use machine learning and a large national dataset to develop a more personalised estimate of cancer prognosis.

Following recommendations from cancer experts, the team collected relevant information from

patients diagnosed in 2015 and 2017 with breast, thyroid, and pancreatic cancers from the National Cancer Database – a database containing records for 72% of newly diagnosed cancer cases in the US. The team selected breast, thyroid, and pancreatic cancers as 'proof of concept' sites, based on their diversity in terms of the composition of patient populations and survivability.

Following the first phase, three-quarters of the collected data were used to train the machine learning algorithms to recognise patterns between patient characteristics at diagnosis and patient survival at five years, and then to rank the factors with the greatest influence on survival. The remaining quarter of the collected data was reserved to test the prediction model's accuracy in estimating patient survival.

Altogether, the team included data from 259,485 breast cancer patients, 76,624 thyroid cancer patients, and 84,514 pancreatic cancer patients.

Results from random forest algorithms, ranking the most highly influential variables in terms of patient survival, found that the top four factors influencing whether patients were alive five years after diagnosis according to cancer site were:

- For breast cancer: whether the patient had surgery, the patient's age at diagnosis, tumour size and time from diagnosis to treatment.
- For thyroid cancer: age at diagnosis, tumour size, time to treatment, and lymph node involvement.
- For pancreatic cancer: cancer surgery, histology or microscopic analysis of cancer, tumour size and age at diagnosis.

Additionally, the team found that hormone receptor status and the presence of ki67 biomarkers were significant survival factors for breast cancer patients.

"Overall, we have demonstrated proof of concept of the ability of our cancer survival calculator to take into consideration patient treatment, and tumour specific factors, in addition to tumour stage," says Janczewski.

For the individual patient, the final product from the process will be the generation of an estimated survival curve showing survival probability (or percentage chance of being alive) across the course of follow-up. "The prototype is meant to be used at the initial clinical encounter, but we plan to make this more dynamic, so that it can be used at any point in time," says Janczewski.

The next step, she adds, would involve developing a 'web application' that would be accessible by clinicians for use in clinical practice and conducting pilot tests at selected cancer centres. Plans include generating models for all cancers in the National Cancer Database including colon, lung, hepatobiliary, skin (melanoma), and prostate cancers. "It will be important for patients to be appropriately supported and counselled during the process," says Janczewski.