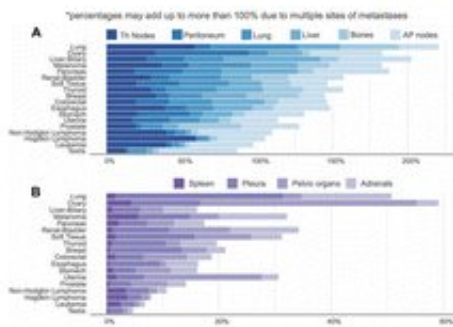


# Predicting cancer spread with natural language processing

November 18 2021

## Patterns of Metastatic Disease in Patients with Cancer Derived from Natural Language Processing of Structured CT Radiology Reports over a 10-year Period



Graphs show (A) common sites of metastatic disease and (B) less frequent sites of disease.

- Natural language processing can achieve high accuracy (90%–99%) in extracting metastatic disease labels in different organs within structured radiology reports for patients with cancer.
- In chest, abdomen, pelvis CT, abdominopelvic nodes are the sites most frequently reported to have metastatic disease (23.6% of all reports), followed by thoracic nodes (17.6%), lungs (14.7%), liver (13.7%), and bones (9.9%).
- Metastatic disease patterns of spread are distinct among common cancers, with bone affected first most frequently among breast and prostate cancers, liver among colorectal and pancreas cancers, and lung among urinary and lung cancers.

Do RKG et al. Published Online: July 27, 2021  
<https://doi.org/10.1148/radiol.2021210043>

Radiology

Credit: DOI: 10.1148/radiol.2021210043

Gathering data from CT scans can be labor intensive and exhaust health care resources. Queen's researchers Amber Simpson and Farhana Zulkernine along with radiologist Richard Do (Memorial Sloan Kettering Cancer Center, New York) are developing technology that will relieve these issues, as well as predict how cancer will spread in patients, using natural language processing.

Natural language processing (NLP) is used to program computers to

process and analyze large amounts of language data from interactions between humans and computers. Dr. Simpson (School of Computing; Biomedical and Molecular Sciences) and Dr. Zulkernine (School of Computing) have leveraged the data scraping abilities of NLP, applying the technology to CT scans to predict where cancer could spread.

"Artificial Intelligence (AI) has the potential to solve fundamental problems in cancer that cannot be solved by humans alone," says Dr. Simpson, who is the Canada Research Chair in Biomedical Computing and Informatics. "For example, we do not have knowledge of how chemotherapy behaves in the general population of cancer patients. Chemotherapy is tested in clinical trials with strict criteria on the patients to include and exclude. AI gives us the opportunity to study cancer response and spread across the entire cancer population."

To develop NLP models, three radiologists curated a sample of over 2,200 CT scans, searching across 13 organs for the presence or absence of metastatic cancer. Three different models were then tested with nearly 400,000 CT scans, with the best performing model reaching rates of 90–99 percent accuracy for detecting and labeling cancer across all organs. These results were published in the journal *Radiology*.

"The radiology reports we had access to had only semi-structured text data," said Dr. Zulkernine, explaining why they decided to take an interdisciplinary approach to the research using NLP. "Therefore, with School of Computing students Karen Batch and Kaelan Lupton, we developed an NLP pipeline to pre-process the data and extract key features to feed into a machine learning model to make predictions about cancer metastases in 13 different organs based on the prior reports."

Using AI to detect cancer metastasis means that knowledge of every cancer patient, not only those on trials, can be brought to bear for

individual patient diagnoses and treatment plans.

But the potential doesn't stop there. With the analysis provided from the 400,000 CT scans, Dr. Simpson and Dr. Zulkernine plan to create a digital cancer twin that would mirror and predict the spread of cancer in a patient. Knowing where and how cancer could spread means localized, not systemic, treatment.

"Precision medicine is one of the greatest opportunities and obstacles in modern cancer care," says Dr. Zulkernine. "A targeted, localized therapy is definitely better as metastasis in different organs may be treated with more focus on the type and severity of the metastasis, without affecting the complete biological system or other organs."

The concept of a digital twin comes from the manufacturing field, where it is used to improve processes by monitoring and producing a constant feedback loop. When applied to treating patients, the digital twin could collect and analyze data that would improve patient's health and quality of life. This work, which is funded by the New Frontiers in Research Fund program, is being done in collaboration with Sharday Mosurinjohn (School of Religion).

The use of AI, particularly in health care, raises many philosophical questions, as well as social concerns. Dr. Mosurinjohn will perform contextual analysis alongside the development of this technology to further investigate the possible ramifications of the digital twin. Factoring in such concerns will allow for [cancer](#) patients to have a more meaningful role within their own treatments.

**More information:** Richard K. G. Do et al, Patterns of Metastatic Disease in Patients with Cancer Derived from Natural Language Processing of Structured CT Radiology Reports over a 10-year Period, *Radiology* (2021). [DOI: 10.1148/radiol.2021210043](https://doi.org/10.1148/radiol.2021210043)

Provided by Queen's University

Citation: Predicting cancer spread with natural language processing (2021, November 18)  
retrieved 30 June 2024 from <https://medicalxpress.com/news/2021-11-cancer-natural-language.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.