The future of oncology: Using machine learning to drive the development of personalized medicine

Erica Brant (BScH)[1]

[1]Department of Medicine, University of Toronto, Medical Sciences Building, 1 King's College Circle, Toronto, ON, Canada, M5S 1A8.

Abstract

In recent years, healthcare has shifted towards treating each cancer patient as an individual, rather than one of thousands with a similar disease. With the advances that have been made in molecular profiling, it is no longer feasible to apply a broad cancer treatment when there is a plethora of tumour variability. In this day and age, there is a vast amount of healthcare data at our fingertips which can be targeted and analyzed using machine learning. Machine learning, a subset of artificial intelligence, allows physicians to quickly identify evidenced-based personalized treatments specific to each patient and the molecular profile of their particular tumour. In addition to identifying the optimal treatment dosage to maximize efficacy and minimize side effects, machine learning can standardize information received by patients regardless of their physical location and the physician that they see. Although machine learning has the potential to change how we treat cancer, research is a long way from actualizing this goal. With time, research, and investments, our colleagues of the future will be utilizing this technology in their everyday practice to tailor treatments to each individual, rather than offering the same blanket treatment to all.

It is every clinician's dream that cancer will one day be a disease of the past. The field of medicine aspires to reach a point where early diagnosis, improved quality of life, and a cure for cancer is not a far-flung hope. Despite the attractiveness of this idea, cancer remains a part of our lives and medical practices, with many patients unfortunately who do not achieve successful remission. In the age of technology where data sharing is possible, it is time to approach cancer treatment through a different lens. It is no secret that healthcare as a whole is striving to move towards ensuring that each patient remains an individual, with their own personal treatments as opposed to a cookie-cutter approach. This is the principle behind personalized medicine. It takes the unique qualities of a patient's disease and recommends a treatment specially tailored to their disease characteristics. It has the ability to offer an estimate of overall prognosis, as well as a probability of recurrence. Machine learning (ML) is a new and promising weapon in "the War on Cancer", and has the potential to assist physicians in identifying ways to personalize treatments for each patient and their unique tumour genetics.[1]

Before delving into how ML can be utilized in cancer care, it is important to first understand why personalized medicine is needed for cancer treatment. This can be accomplished by appreciating the complexity of the disease itself. As we know, cancer is not one disease, but rather a myriad of diseases differing in location, cell of origin, and the mutations that promote oncogenesis.[2] There is interpatient heterogeneity to consider, which involves genetic differences in tumours between different patients.[3] In addition, both spatial and temporal intratumour heterogeneity must also be considered. Spatial intratumour heterogeneity refers to genetic alterations that different subpopulations of cells have in the same tumour geographical region, whereas temporal intratumour heterogeneity speaks of the genetic alterations that exist as the disease progresses and reacts to selective pressures.[3] An example of a consequence of temporal heterogeneity is in the setting of metastasis, and the acquisition of alterations that allow prostate cancer to enter bone, for instance. This heterogeneity complicates current tumour sampling techniques, needle biopsies and surgical excision, making it unlikely to attain a complete profile of a patient's cancer using these methods.[3] For example, core biopsy tissue samples used to diagnose prostate cancer can vary according to inter-and intratumour heterogeneity.[4] This could be detrimental to a prostate cancer patient who might have an observed Gleason score lower than it actually is, causing them to possibly miss out on early life-saving interventions. With all of these factors to consider, there is no one treatment for cancer that is equipped to handle this degree of heterogeneity at this time. However, a promising starting point lies in genetic alterations.

Conventionally, cancer has been treated in accordance to what organ is being affected, or where the cancer has originated from.[5] Considering the vast literature on the heterogeneity of cancer, we are now able to explain why patients with seemingly similar cancers can respond to treatment in an equally heterogeneous way, with some successfully treated and others nonresponsive.[5] This "one-size-fits-all" approach is clearly not effective, and the idea of personalized medicine is looking to fill the treatment gap. Personalized medicine means being able to choose a treatment regimen based on specific genetic alterations that are present in

Corresponding Author:
Erica Brant
erica.brant@mail.utoronto.ca

a patient's cancer. In addition, the goal of personalized medicine is to be able to prescribe the correct dose of the ideal drug (i.e., a targeted therapy), in order to eliminate or minimize side effects as much as possible.[6]

The concept of personalized medicine combines information gained from a patient's tumour, cell-free DNA, immune markers, and takes comorbidities into account in order to propose the optimal therapy for that patient.[7] Cancer is a disease of the genome; by identifying the differences in gene expression and unique cell characteristics, new targeted treatments could be researched and actualized. With the ability to sequence entire cancer genomes, and with initiatives such as the Cancer Genome Project collecting said data, targeted treatments are starting to be used based on research findings regarding altered molecular pathways in cancer.[8] For example, trastuzumab is used for HER2/neu positive breast cancers, while imatinib is used in chronic myeloid leukemia to target BCR/ABL tyrosine kinase.[9,10] The standard treatments are not effective for all patients, and in some cases, may actually harm more than help. For example, one would not treat a triple-negative breast cancer patient with trastuzumab, as they lack HER2 overexpression.[11] The wealth of knowledge around molecular pathways has garnered attention from pharmaceutical companies, and many new cancer drugs targeting these molecular pathways have been released. So far, these individual targeted treatments have been efficacious for some patients, but there have been heterogeneous responses in patient populations, even with a similar mutation. For instance, in EGFR-mutant non-small cell lung cancer, the targeted EGFR therapy is ineffective in around 30% of patients.[12]

With over 100 types of cancer, the power to be able to give a precise, successful treatment is as daunting as it is tantalizing. Though there have been small victories in this department to date, it is not perfect. The current targeted therapies are far from being a cure. Consider, EGFR-targeted therapies for EGFR-mutant non-small cell lung cancer merely shift the median overall survival from 8-12 months to 20-30 months.[13] In order to develop the prospect of personalized medicine further, identify more target pathways or mutations, or offer unique combinations of drugs to patients based on many unique cancer characteristics, a drastic change must be made. Indeed, the way to true success will be paved through the use of artificial intelligence (AI), ML, and the extensive amount of healthcare data available.

In medicine, AI is a branch of computer science that has the ability to analyze large sets of medical data.[14] The premise regarding AI is to have a computer system function with characteristics of human intelligence, and in medicine, to aid the physician in formulating a diagnosis and appropriate treatment plan, as well as offer a prognosis for the patient's disease.[14] ML (a subset of AI) refers to the application of algorithms that access large data sets to allow the computer to learn through "experience" with data and results, as a human would.[15] Broadly speaking, ML provides a program with a baseline of data and its conclusion – for example, the presence of a KRAS gene mutation and the likelihood of malignancy in colon cancer - to allow it to analyze future data sets for similar patterns. At its peak, a physician could pose the question "is this tumour malignant?" to the program, and after sorting through and examining health data, it could offer the probability of malignancy.[16] Furthermore, ML could be used to predict cancer susceptibility, cancer recurrence, and extract other patterns from healthcare data through data mining. It is the latter application that would allow clinicians to observe new patterns between cancer subtypes that have not been known before, and offer new targets for future treatment.

To delve into creating a standard of personalized care, we need the ability to examine and draw conclusions from "Big Data." However, it simply is not feasible to expect any clinician or researcher to sift through this sea of data in a timely matter.[17] By applying algorithms to the data, ML would be able to quickly sort through and draw conclusions faster than healthcare professionals could ever dream of doing on their own time.[18] It is laborious and time-consuming to be completely up to date with medical literature where there are ever evolving guidelines, and new drugs being approved. U.S. physicians, for example, only have about 4.6 hours a week on average to obtain new knowledge.[19] The fact remains, the data that we have at our fingertips is too large for a human to handle, and our machine counterparts have become the only tool that can do this job efficiently.

This application of ML will allow for evidence-based medicine to reach the practicing oncologist and primary care physician in a timely manner, and allow for the development of a targeted and effective treatment to a patient's particular subtype of cancer.[18] ML has the potential to improve how we deliver safe and effective care, as well as standardize the information a patient receives regardless of where they live and what doctor they see.[20] This technology will become a tool in our arsenal that we can use to help to better guide our clinical decisions.

What does this all mean? ML for personalized medicine has potential, but it is still in its infancy. We have started to make strides in the way that we target markers in individual cancers and we have subsets of populations that are responding remarkably to some of the targeted treatments that we have. Not only could ML help identify the correct treatment, but it could be applied to suggest doses of the treatment to maximize efficacy and minimize side effects.[21] It could highlight treatments most likely to be useful, second-line treatments, or treatments that would not be recommended for a patient. For the ease of the physician, the knowledge would be easily accessible. All that is needed is to harness our skills of interpretation and human understanding to carry it through to our patients. For pharmaceutical companies, this may seem limiting at first, but these algorithms have the potential to uncover similar molecular profiles between tumours that originally were assumed to be vastly different, thus possibly paving the way for the current targeted therapies to be utilized for more patients.[5]

There are companies that are attempting to utilize this technology for personalized medicine. Oncotype DX, by Genomic Health, is one of these technologies currently on the market. One such application of detailed data analysis, the Oncotype DX Breast DCIS Score test, provides physicians with a patient's individualized assessment of recurrence for breast cancer - specifically ductal carcinoma in situ - by examining 21 genes.[22] Even though this test can help to influence treatment decisions, it only offers a useful risk assessment in 60% of patients, and it can place patients in only an intermediate risk category when pathology suggests a higher risk.[23] Information from tests such as Oncotype DX has the potential to be useful, and has already influenced therapy, but development needs to continue in order to fully realize the potential of ML.

ML is as only as effective as the data that we possess. There is variation among professionals, contradictory work, inconsequential studies with non-causal work, and so on, that influence the data that we have.[17] It is no easy feat to identify relevant work among the millions of entries in "Big Data". Resultantly, we run the risk of harm to our patients if we do not approach the use of personal health data with the appropriate precautions. For example, it could lead to misleading interpretations of ML conclusions for targeted therapies, thus affecting patient care and having potential negative consequences on the success of overall treatment.[24] Even with the genetic mutations that we have identified so far, it is not possible to create therapies targeted to all of them. It is estimated that only 9% of genetic mutations are actionable at this point.[25] Furthermore, not every patient will respond the same to the targeted therapies, as tumour heterogeneity is still an issue with personalized medicine at this time. With further research, development, resource-sharing, training, and investment, we will be able to offer patients their individualized treatment.

In order to see healthcare and AI in clinical practice, there are many hurdles that need to be overcome, and regulations that need to be made. Guidelines need to be created for the use of ML to analyze data. In terms of the data itself, there are many factors that need to be considered and resolved. Language and culture will be a barrier to sharing and understanding electronic records, as well as personal data privacy.[26] Of course, a number of questions remain. Who owns the data? Will companies be willing to share data in order to further cancer care? It is critical to ensure that the data is protected, and kept from misuse. As clinicians embark on the journey to move towards personalized medicine using ML, we can start by continuing to identify patterns in cancer subtypes to identify possible targets for future therapy, genes that convey a risk for recurrence, or to give a prognosis.

We have every reason to be optimistic about the role that ML will play in the future of cancer care, and the further development of personalized medicine. Our future medical students, physicians, and other health professionals will likely see the fruits of ML in their daily lives, thus it is critical for us as health professionals to understand the impact that ML has now and what it could have down the road. We have already seen hints of this with our current existing targeted therapies, and companies being able to offer an individualized risk assessment, depending on the cancer type. It is exhilarating to imagine a day when we are able to pinpoint molecular profiles in each individual patient's tumour and apply ML to identify accurate, successful, evidence-based treatments and dosing. We can no longer use the "one-size-fits-all" approach from the days of treatment past. Our success with cancer treatment will be made through recognizing the individual, rather than treating the identified broad and variable cancer type. With ML at its helm, personalized medicine is lining up to be one of the most revolutionary advances in cancer care.

References

1. National Cancer Institute (NCI) [Internet]. NCI:National Institutes of Health. National Cancer Act of 1971; 2016 [updated 2016 Feb 16; cited 2018 Dec 31]. Available from: https://www.cancer.gov/about-nci/legislative/history/national-cancer-act-1971.
2. Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas Pan-Cancer Analysis Project. Nat Genet. 2013;45(10):1113-1120: doi: 10.1038/ng.2764.
3. Bedard PL, Hanson AR, Ratain MJ, et al. Tumour heterogeneity in the clinic. Nature. 2013;501(7467):355-364: doi: 10.1038/nature12627.
4. Cicione A, Cantiello F, De Nunzo C, et al. Needle biopsy size and pathological Gleason Score diagnosis: No evidence for a link. Can Urol Assoc J. 2013;7(9-10): E567–E571: doi: 10.5489/cuaj.311.
5. van 't Veer L, and Bernards R. Enabling personalized care medicine through analysis of gene-expression patterns. Nature. 2008;452:564-570:doi:10.1038/nature06915
6. Verma M. Personalized medicine and cancer. J Pers Med. 2012; 2(1): 1–14: doi: 10.3390/jpm2010001
7. Fountzilas E, Tsimberidou AM. Overview of precision oncology trials: challenges and opportunities. Expert Rev Clin Pharmacol. 2018;11(8):797-804.
8. Tannock IF, Hickman JA. Limits to personalized cancer medicine. N Engl J Med. 2016; 375(13):1289-94.
9. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med. 2001;344(11):783–92.
10. Druker BJ, Sawyers CL, Kantarjian H, et al. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. N Engl J Med. 2001;344(14):1038–42.
11. Wahba HA, El-Hadaad H. Current approaches in treatment of triple-negative breast cancer. Cancer Biol Med. 2015;12(2):106-16.
12. Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. N Engl J Med. 2010;362(25):2380–88.
13. Mitsudomi T, Morita S, Yatabe Y, et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol. 2010;11:121–128.
14. Ramesh AN, Kambhampati C, Monston JRT, et al. Artificial intelligence in medicine. Ann R Coll Surg Engl. 2004;86:334-338: doi 10.1308/147870804290.
15. Holzinger A, Dehmer M, and Jurisica I. Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions. BMC Bioinformatics. 2014;15(6):I1: doi: 10.1186/1471-2105-15-S6-I1.
16. Kourou K, Exarchos TP, Exarchos K, et al. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8-17.
17. Kohn MS, Sun J, Knoop S, et al. IBM's health analytics and clinical decision support. Yearb Med Inform. 2014;9(1):154-62.
18. Liu C, Liu X, Wu F, et al. Using artificial intelligence (Watson for oncology) for treatment recommendations amongst Chinese patients with lung cancer: feasibility study. J Med Internet Res. 2018;20(9):doi:10.2196/11087.
19. Woolhandler S, Himmelstein DU. Administrative work consumes one-sixth of U.S.physicians' working hours and lowers their career satisfaction. Int J Health Serv. 2014;44(4):635-42.
20. Golden JA. Deep learning algorithms for detection of lymph node metastases from breast cancer: helping artificial intelligence be seen. JAMA. 2017;318(22):2184-86.
21. Clark PM, Ebiana VA, Gosa L, et al. Harnessing preclinical molecular imaging to inform advances in personalized cancer medicine. J Nucl Med. 2017;58(5):689-96.
22. Oncotype IQ [Internet]. About Oncotype DX DCIS Score. Oncotype IQ: Genomic Health; 2018 [cited 2018 Dec 31]. Available from: https://www.oncotypeiq.com/en-US/breast-cancer/healthcare-professionals/oncotype-dx-breast-dcis-score/about-the-test.
23. Kelly CM, Krishnamurthy S, Bianchini G, et al. Utility of oncotype DX risk estimates in clinically intermediate risk hormone receptor-positive, HER2-normal, grade II, lymph node-negative breast cancers. Cancer. 2010;116(22):5161-5167: doi: 10.1002/cncr.25269.
24. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517-18.
25. Prasad V. Perspective: The precision-oncology illusion. Nature. 2016;537(7619): doi: 10.1038/537S63a.
26. Rabbani M, Kanevsky J, Kafi K, et al. Role of artificial intelligence in the care of patients with nonsmall cell lung cancer. Eur J Clin Invest. 2018;48(4): doi: 10.1111/eci.12901.