Patient centered care for prostate cancer—how can artificial intelligence and machine learning help make the right decision for the right patient?

Nathan C. Wong, Bobby Shayegan

Division of Urology, Department of Surgery, McMaster University, Hamilton, Ontario, Canada

Correspondence to: Bobby Shayegan, MD, FRCSC. 50 Charlton Ave. E., Hamilton, ON, Canada. Email: shayeb@mcmaster.ca.

Provenance: This is an invited article commissioned by the Section Editor Peng Zhang (Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, China).

Comment on: Auffenberg GB, Ghani KR, Ramani S, *et al.* askMUSIC: Leveraging a Clinical Registry to Develop a New Machine Learning Model to Inform Patients of Prostate Cancer Treatments Chosen by Similar Men. Eur Urol 2018. [Epub ahead of print].

Submitted Dec 21, 2018. Accepted for publication Jan 03, 2019.
doi: 10.21037/atm.2019.01.13
View this article at: http://dx.doi.org/10.21037/atm.2019.01.13

In the era of advancing technologies, large volumes of data are being collected by electronic medical records and clinical registries are readily available for data mining (1). These registries, however, require appropriate analyses and interpretation to derive clinically meaningful benefits for patients. Traditional statistical models have been previously used for this purpose. However, they are incapable of processing highly dimensional data and do not actively adapt with the incorporation of new data points. Although information from clinical registries assist physicians in making data-driven decisions, there are limited opportunities for patients to directly interact with these registries to help them make informed decisions.

askMUSIC is a novel web-based system that utilizes a prospective prostate cancer registry, the Michigan Urological Surgery Improvement Collaborative (MUSIC), to provide men newly diagnosed with localized prostate cancer with an opportunity to view treatment decisions of patients with similar characteristics to themselves have previously made. Auffenberg *et al.* utilized this data to train a model to predict patient treatment decision using the random forest machine learning method (2). The registry used for this included 7,543 men diagnosed with prostate cancer, of whom 45% were treated with radical prostatectomy, 30% surveillance, 17% radiation, 6% androgen deprivation therapy, and 2% watchful waiting. The data was subsequently divided into training and validation cohorts using a 2:1 random split stratified by practice location. Overall, the personalized model was highly accurate with an area under the curve of 0.81 and age, followed by number of positive cores and Gleason score, appeared to be the most influential variables that influenced patient treatment decisions.

This convenient tool is readily available online for direct patient use (https://ask.musicurology.com/). The website offers a wide range of prostate cancer information from risk factors to the description of treatment options. It is simple, elegant and patient-friendly. After patients input variables including their age, prostate specific antigen (PSA), Gleason score (primary and secondary), number of positive cores, total number of cores, patient weight and a limited comorbidity estimate (history of myocardial infarction and diabetes), the machine learning algorithm provides them with the answer to the question: "what treatments did patients similar to me choose?".

Machine learning is not new and has been increasingly applied to various fields from economics to medicine. Developed from computer sciences and a subset of artificial intelligence, it semi-automatically trains algorithms that are able to uncover complex patterns and non-linear relationships faster and more accurately than conventional statistical models. By evaluating kinematic and events data, machine learning algorithms have been used to develop and validate performance metrics of surgeons performing robot-assisted radical prostatectomy, laying the foundation for standardized metrics in the evaluation, training and quality improvement (3). From histological, clinical and demographic information, machine learning algorithms can also directly assist in patient care by improving accuracy of the prediction of disease outcomes and prognosis (4). Various machine learning algorithms have been shown to outperform traditional statistical models for predicting early biochemical recurrence following robot-assisted radical prostatectomy (5). With the incorporation of larger volumes of data generated from genomic profiling, predictive algorithms can be used to individualize and personalize treatments for target patients. Artificial intelligence has the power to sift through the vast amounts of complex variables and identify highly complex interactions.

The treatment decision process for men newly diagnosed with prostate cancer is complicated and demands a shared decision-making process as there are various treatment options available, including surveillance. The PROTECT trial randomized 1,643 men to active monitoring, radical prostatectomy and radiotherapy (6). At 10 years median follow-up, cancer specific mortality was low regardless of the randomized treatment and all treatments were statistically similar. Although surgery and radiation were both associated with a lower incidence of cancer progression and metastases, they were also associated with more complications that negatively affect patients' quality of life. An informed decision is a complex decision-making process based on various factors including the natural history of the disease, outcomes of treatment choices, adverse effects and predominantly patient preferences dependent on patient-specific variables.

Machine Learning based on clinical registry data is a novel strategy for clinical decision-making. askMUSIC allows patients to directly interact with the model and feel in control of their disease, treatment and life. Empowered with objective information, this tool may enhance the patient decision-making process and increase satisfaction of care. However, the appropriate application of machine learning algorithms is dependent on 3 fundamental elements: the input variables, the predictive outcome and the machine learning model used.

The quantity and, more importantly, the quality of input variables have an enormous impact on the performance and generalizability of machine learning models. Complete data is typically required and missing data points are either completely excluded or are estimated based on non-missing data. Although a good proportion were excluded by date of diagnosis to build a contemporary cohort, the MUSIC registry had 29,862 patients but the final study cohort which the machine learning algorithm is based on included only 25% of the entire registry. The input variables were also limited to 8 patient factors, likely to make the online model more user friendly. However, machine learning has the ability to process large volumes of data in a hypothesis free-manner and other factors including family history and clinical stage which are commonly included in prostate cancer nomograms, can be included to improve predictive potential.

The predicted outcome of patient treatment decision is dependent on the patient population in which the algorithm is trained on. These treatment decisions are likely biased not only by what the treating physician offers, but also by patient characteristics including personality, values and culture. Although MUSIC is a collaboration between 45 diverse community and academic centers representing the majority of urologists who practice within Michigan, the predicted outcomes are likely different compared to other regions around the world. Thus, these algorithms may not be directly applicable to all patients and require validation. Furthermore, treatments will continue to evolve over time with a higher utilization of active surveillance and with the evolution of focal therapy. Regardless, the appeal of machine learning is that the models continue to learn and are easy to be "retrained" with the incorporation of new data, maximizing predictive accuracy in different patient populations with diverse demographic and clinical profiles.

Finally, there are several machine learning models available, each with their pros and cons. The publication by Auffenberg *et al.* reported only on the random forest machine learning model. This model utilizes multiple decision trees and uses random combinations of independent variables to predict the outcome of interest. Using a majority vote system, a new sample or "patient" is predicted using these decision trees, and the ultimate classification of the new label or "patient outcome" is determined by the prediction reached by the majority of the trees. Other machine learning models exist including, but not limited, to K nearest neighbor, support vector machines, Bayesian network, artificial neural network and more. Training the correct model for the appropriate application can be a challenging task. For example, the K nearest neighbor algorithm a simple model that is useful in large volume classification and when data is irregular. However, it is highly sensitive to variability of local data points compared to the overall pattern. The basic principle is to use a predefined number of training data points, also known as "K", with known labels that most closely resembles the new data point, and thus predicting the new label from the “K-nearest” training points.

With the incorporation of further data points and model retraining, we expect the predictive accuracy of machine learning to continue to improve and provide personalized medicine (1). Larger patient datasets and electronic medical records can be surveilled in a semi-automated manner to providing instantaneous predictive analytics that can be used to derive insights into a variety of diseases. However, the prediction accuracy depends heavily on effective data integration acquired from various sources to allow it to be generalized. Although these models will not replace the share decision-making process, it may compliment the information that patients receive from traditional educational models. While this is only the beginning and further validation is needed, the future application and possibilities of machine learning and artificial intelligence are endless.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Cite this article as: Wong NC, Shayegan B. Patient centered care for prostate cancer—how can artificial intelligence and machine learning help make the right decision for the right patient? Ann Transl Med 2019;7(Suppl 1):S1. doi: 10.21037/atm.2019.01.13

References

1. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016;375:1216-9.
2. Auffenberg GB, Ghani KR, Ramani S, et al. askMUSIC: Leveraging a Clinical Registry to Develop a New Machine Learning Model to Inform Patients of Prostate Cancer Treatments Chosen by Similar Men. Eur Urol 2018. [Epub ahead of print].
3. Hung AJ, Chen J, Jarc A, et al. Development and Validation of Objective Performance Metrics for Robot-Assisted Radical Prostatectomy: A Pilot Study. J Urol 2018;199:296-304.
4. Kourou K, Exarchos TP, Exarchos KP, et al. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17.
5. Wong NC, Lam C, Patterson L, et al. Use of machine learning to predict early biochemical recurrence after robot-assisted prostatectomy. BJU Int 2019;123:51-7.
6. Hamdy FC, Donovan JL, Lane JA, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med 2016;375:1415-24.