Using Big Data Analytics to Advance Precision Radiation Oncology

Todd R. McNutt, Stanley H. Benedict, Daniel A. Low, Kevin Moore, Ilya Shpitser, Wei Jiang, Pranav Lakshminarayanan, Zhi Cheng, Peijin Han, Xuan Hui, Minoru Nakatsugawa, Junghoon Lee, Joseph A. Moore, Scott P. Robertson, Veeraj Shah, Russ Taylor, Harry Quon, John Wong, Theodore DeWeese

Our Project: Anomaly detection for treatment planning and a learning health system in radiotherapy

- Members: Vincent Qi, Daniel Yuan
- Mentors: Todd McNutt, Pranav Lakshminarayanan

Learning health system(LHS)

- A learning health system (LHS) is a concept where quantifiable diagnostic, treatment and outcome data are captured from a continuous stream of patients and placed in a knowledge base.

- Knowledge is accessed by analytical tools that employ statistical and machine learning algorithms to present trends and make predictions and causal inferences on outcomes

Introduction

▶ The goal of precision medicine is to improve overall patient care and determine when and how to personalize patients' treatments. Currently, this is guided by a physician's understanding of the patient's condition by drawing from their experience to align the specifics of care to the patient.

▶ Guidelines assist in the overall pathways for specific diseases, but, for the most part, precision medicine is performed with finer granularity than the guidelines provide.

Decision making

▶ In decision making, we decide on the most appropriate intervention for the patient which may or may not be guided by complete knowledge of the underlying biological mechanisms. New discovery however, must uncover the biological understanding or derive hypotheses that may be further validated under more controlled studies.

Big Clinical Data

▶ The ability of big clinical data, to represent the real world with minimal bias, to accumulate assessments over time, to be linked with other databases, to be used and reused, and to enable a multidimensional understanding, should all be considered to unlock the potential

▶ Clinical data represent prior experience from patients and are captured through a multitude of methods, but limitations of our current protocols and pathways result in only a small fraction being used to make clinical decisions

▶ Clinical Data generally have a number of complications not found in typical cross-sectional study datasets. For example, clinical data exists in forms of free text to three-dimensional volumes to structured data elements all with longitudinal sampling. Clinical data also suffer from selective sampling, missingness, and measurement error.

Problems with big data

- Aside from lifestyle covariates, clinical data contains patient and disease status, treatment and symptom management, clinical and quality of life (QoL) outcomes, adverse effects, and survival. The key for enabling access is to extract meaningful information or features and store them in standardized ways

- the level of precision in measuring outcomes dictates the quality of subsequent clinical conclusions

- The measurement of a patient's clinical condition depends on available time and resources

- Longitudinal assessment of patient status requires careful feature extraction.

- Example, taste disturbance during treatment to understand coping, or evaluating longer term toxicities to provide a measure of permanent damage.

Problems with big data

- Unlike standard cross-sectional studies, where treatments are binary and represent case and controls groups, radiation therapy involves a three-dimensional dose delivered over multiple days, yet protocol standards extract simplistic dose-volume features as efficient measures of treatment plan quality

- Dose Volume Histograms (DVHs) leave out useful information, and thus are insufficient on their own to support precision medicine.

- A DHV assumes each location within a region is equally sensitive to radiation and equally responsible for biological function.

# The learning health system and predictive modeling

▶ A common goal of traditional statistical modeling is the discovery of the underlying mechanisms or cause of outcomes.

▶ the "data models" are usually hypothesis-driven, yet may not reflect the complexity of the true process, but nonetheless enable improved understanding of the system. The "algorithmic models", on the other hand are hypothesis-generating presenting superior predictive accuracy, yet make it challenging to uncover the dominant input variables and/or causal attributes.

- Medical information is very complex and often aggregated into features that can mask important underlying details

- Such dimension reductions are necessary, but risk being insufficient

- This data reduction may have a negative impact on the ability to build a model to predict organ function or disease control after treatment that may have spatial dependence

- It is not easy to proactively determine whether this type of ad hoc feature will preserve or discard useful relationships between the features and outcomes.

- Considerations for predictive models must include the purpose of building them, whether they are to be used for decision support or for discovery of new knowledge. There is more than one tool and selecting the right one to apply to the clinical question and purpose will be critical for making more precise patient care decisions.

Decision support

- The goal of decision support is to provide the most appropriate intervention for the patient and not necessarily to discover new knowledge. This begs the question of which outcome prediction models should be selected with what accuracy requirement

- The key to selecting the more performative model is understanding the decision and intervention to be made

- Example: feeding tube to treat weight loss during radiotherapy, or modify treatment to prevent taste disturbance
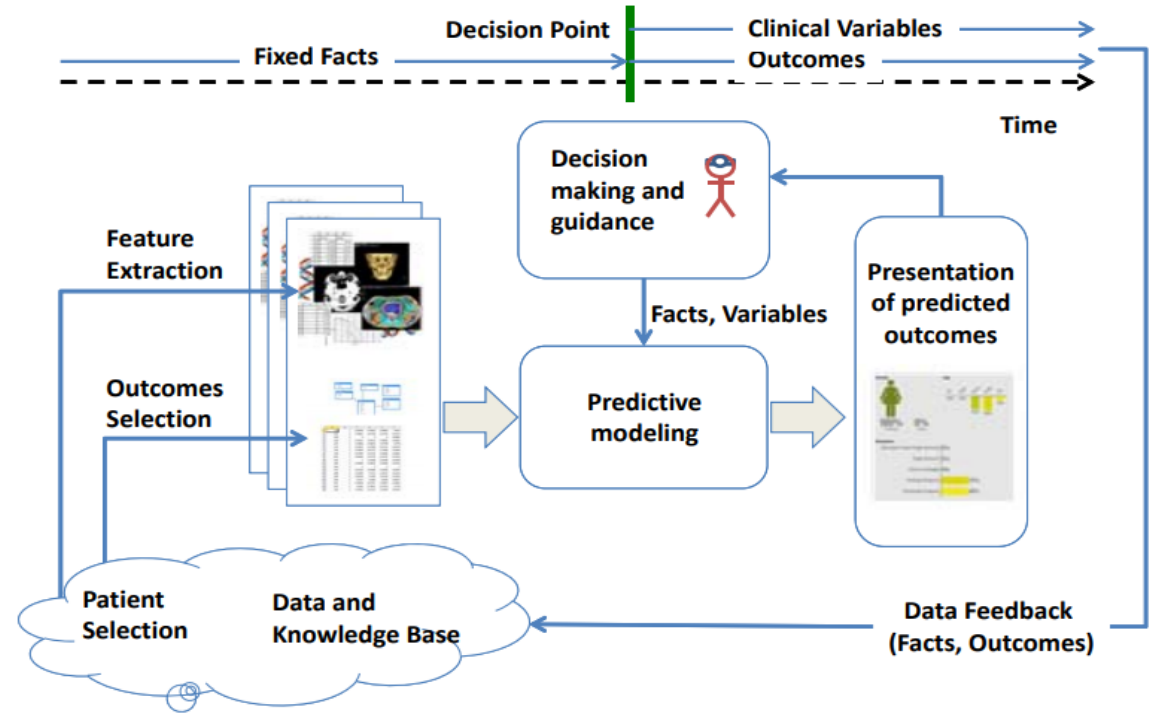
Figure 1: Decision support framework to make predictions of outcomes of individual patients. The models, derived from the knowledge base, use the facts and clinical options/variables in making the predictions which are presented to the physician to assist in decision making.

Discovery and hypothesis derivation

▶ A LHS also provides the opportunity to extend knowledge through discovery and hypothesis derivation

▶ In essence, the goal is to both understand features most predictive of outcomes and uncover the underlying causes
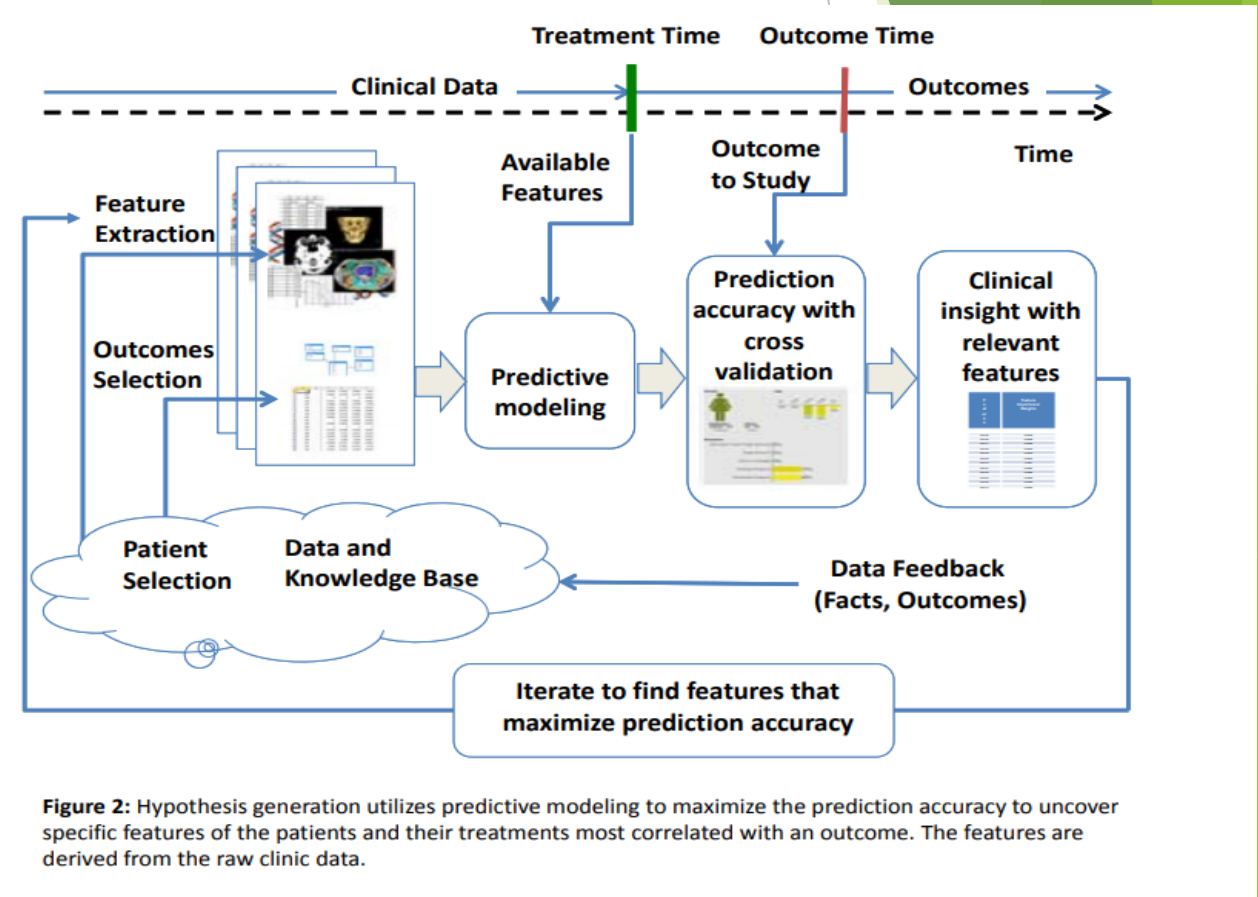
Figure 2: Hypothesis generation utilizes predictive modeling to maximize the prediction accuracy to uncover specific features of the patients and their treatments most correlated with an outcome. The features are derived from the raw clinic data.

Discovery and hypothesis derivation

- Aside from predictive modeling, cause and effect relationships between features and outcomes are important types of hypothesis and are often the most scientifically relevant

- Though there is a large effort in machine learning and statistics to identify cause and effect relationships from observational data, all causal hypotheses generated by such methods must ultimately be validated by formal randomized controlled trials.

Issues of implementation

▶ Both decision support and discovery are limited by the knowledge contained in the database

▶ This missing of data also manifests itself when models are validated between institutions

Issues of implementation

▶ When using the LHS for decision support the goal is to have the most accurate prediction, and that may happen with models built using only patients treated at the institution where the patient is to be treated

▶ For discovery, however, the goal is to uncover underlying mechanisms, and for this, inter-institutional validation becomes important and completing missing information in the data is crucial to uncovering this new knowledge.

Issues of implementation

- the knowledge base is limited within the norms of clinical care. With radiation treatments, for example, only the variability of the dose distributions present in the knowledge base is available.

Examples

▶ An early example of using big data tools in radiotherapy is the concept of geometry-driven or knowledge-based treatment planning

▶ KBP aligns with the LHS model in that it provides actionable predictions of dose goals for planning and continuously learns as more treatment planning data is accumulated.

▶ In its generalized form, KBP makes use of established machine learning techniques such as supervised inference engines to discover relevant geometric variables and their correlation to patient-specific dose prediction.

▶ The prediction of toxicities is also critical to a patient's ability to tolerate treatment and their long-term QoL. An example is weight loss prediction using a classification and regression tree for head and neck cancer patients

▶ The LHS allows a comprehensive exploration of predictors for a variety of treatment related toxicities beyond the single organ DVH and simple normal tissue complication models, and further, bridging all other clinical and patient factors into an all-encompassing prediction model. Evidence from such models warrant the foundation for clinical decision support for the prevention and/or management of toxicities.

Genomics, pathology and radiomics

- At a higher level, Radiomics, Genomics and Pathology are patient-specific data that are subjected to feature extraction in clinical practice and for research

Their conclusion

▶ Just as machine learning is being used to drive autonomous vehicles, is it reasonable to expect similar successes in radiation oncology? At this point, self-driving cars focus on the rules of the road and respond to immediate detection of obstructions in their local environment

▶ our expectation for the foreseeable future should be one of improved risk/outcome prediction as a supplement to physician-based clinical decision making

▶ The key to success is to uncover and measure as many of the unknowns as possible. Is a future possible in which we accurately measure the critical aspects of patient's outcomes and treatment? Computerization of healthcare is advancing rapidly and the societal culture evolving from having smartphones amplifies the likelihood that good measures of the continuous patient condition will only advance

Their conclusion

▶ As outcome measures improve, radiation oncology must do its part to accurately archive treatments in easily retrievable form, adhering to standard nomenclatures. It should be possible to query features of the patient's history, physical exam, radiographic studies, laboratory tests, and "delivered" dose for any patient from our clinical archive without significant processing. It should be part of the practice to be good stewards of the data and accurately record three-dimensional delivery while capturing the clinical data, appreciating that it ultimately will contribute to the LHS

Evaluation

Pros:

▶ Paper is clear and concise

▶ Argument is balanced discussing issues as well as benefits

Cons:

▶ No concrete discussion about how such a learning health system will be implemented

▶ Future: consider other limitation such as confidentiality and security, legislation or a cost/benefit analysis of creating the framework of such a system.

References

▶ McNutt TR, Benedict SH, Low DA, Moore K, Shpitser I, Jiang W, Lakshminarayanan P, Cheng Z, Han P, Hui X, Nakatsugawa M, Lee J, Moore JA, Robertson SP, Shah V, Taylor R, Quon H, Wong J, DeWeese T, Using Big Data Analytics to Advance Precision Radiation Oncology, *International Journal of Radiation Oncology • Biology • Physics* (2018), doi: 10.1016/j.ijrobp.2018.02.028.