#### Paper Seminar Presentation Project 8: UI for Radiation Therapy Cohort Selection Keefer Chern

#### Paper: Using Big Data Analytics to Advance Precision Radiation Oncology

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.

# **Our Project Goal**

Develop a UI to select for cohorts in a radiation therapy database:

- Use MySQL at the backend to process the medical data
- Use Java to communicate with the backend to extract the data
- JSF (JavaServer Faces) to develop the UI framework
- Allow users to scroll and select desired variables and output patient IDs as cohorts.

### Paper Selection:

- The focus of our project is to extract information from a database based on the parameters given by the user.
- I chose "Using Big Data Analytics to Advance Precision Radiation Oncology" to study the relevance for creating such a UI.
- The goal of our project is a small part of the larger concept of the learning health system this paper develops.

# Background:

- The goal of precision medicine is to improve overall patient care and determine when and how to personalize patients' treatments.
  - Increase quality of care and decrease side effects and cost
- The current method for precision medicine:
  - Heavily dependent on the knowledge the physician possesses for the treatment
  - Intrinsic issues.
  - Guidelines have been established

# Learning Health System (LHS)

- Definition: Concept where quantifiable diagnostics, treatment, and outcome data are captured from a continuous stream of patients and placed in a knowledge base.
  - Can be analyzed with statistical and machine learning tools.
- Comprehensiveness:
  - Support clinical decisions
  - Discovery of important factors
  - Deriving hypotheses

# **Big Clinical Data**

- The LHS needs big clinical data
  - Large database containing clinical information of patients
  - Lifestyle covariates, disease status, symptom management, quality of life outcomes, adverse side effects, survival
- Advantages of big clinical data
  - Can be reused
  - Provide multidimensional understanding
  - Improved by linking with other database

# Usage of LHS

- LHS can provide the useful tool of predictive modeling
- In terms of a data model
  - Describe and validate relationships in data
- Before using a Predictive model
  - Have a reason
  - Choose either Decision Support
  - Or Knowledge Discovery

# **Decision Support**

- Goal: to provide the best intervention for the patient
  - Input: facts and clinical options
  - Output: Possible outcomes
- Key to choose the best model
  - Clear picture of the decision
  - What are possible interventions
- Example 1: deciding whether to use a feeding tube to prevent weight loss for patients
- Example 2: modify toxicity levels to prevent taste disturbance
- Note: Predictive models cannot differentiate between causation from association

### **Decision Support Framework**



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.

# Knowledge Discovery

- Goal: to expand knowledge by understanding what features best predict outcomes and discover underlying causes
- This all helps with deriving hypotheses
  - Cause- and -effect relationships between feature and outcomes
  - Assists decision support
- Still need to be validated through formal controlled trials with real subjects

### Knowledge Discovery Framework



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.

# Data! Biggest limitation of a LHS

- Suffer from selective sampling, missingness, and measurement error
  - $\circ$   $\hfill We suffer from bad data inputting$
- Example 1: Recurrence of Cancer
  - No data on location of recurrence
- Example 2: Method of gathering xerostomia data
  - Scored by a physician
  - Patient questionnaires
  - Controlled stimulation studies

# **Missing Data**

- Example 1: institutions having different systematic care procedures
  - Even worse: procedures not in the database
  - Implication of this cofounder varies
- Example 2: Normalization of Care
  - Unethical to deviate from guidelines just for data
  - Therefore effects of irradiation is hard to explore

#### Data we work with

Any \$Esophagitis: NCI \$Voice Changes 4-D/CT Abs Lymphocyte Count Absolute Lymph Count(JH) Absolute Neut Count(JH) Act Partial Thromboplastin Tim Adenovirus NAT(JH) Admission Reason Admitted Admitted - Y/N AFB Microscopic Exam(JH) AFB Mycobacterial Cult(JH) Agitation - v4 Akathisia - v4 Alanine Amino Trans(JH) Albumin(JH) Alcohol Use (drinks per day) Alcohol Use (years)

VitaminD(25-OH)Total SER(JH) Voice Changes Volume of Radiologic Extra Vomiting - NCI Vomiting - v4 Vomiting CTC4 Watery eye WBC Weight (kg) Weight (lb) Weight Change (kg) Weight Gain - v4 Weight Loss Weight Loss (CTC v4.0) Weight Loss Y/N White Blood Cell Count(JH) White blood cell decreased Wkly Billing Audit Wkly CBCT/EPI Taken & Reviewed Wkly-TX/OTV Scheduled

### **Feature Extraction**

- Radiomics
  - $\circ$  Density, texture, and gradient
- Pathology
  - Cell type, grade level, and differentiation
- Genomics
  - Patterns

# Summary/Key Points

- Learning Health System (LHS) uses big clinical data to produce relevant and useable outcomes.
- Decision support
  - Getting treatment outcome from patient characteristics
- Knowledge discovery
  - Understanding the underlying cause and effect relationships
- Biggest problem
  - Not having well documented and collected data

# Evaluation

#### Pros:

- Very clear about the topic being discussed
- Provides examples for each topic
- Well organized and structured
- Talk about the positives and limitations of each topic

#### Cons:

- Do not go into detail how LHS should be implemented
- No description of the cost of implementing a LHS
- Do not provide solutions to the limitations.
- No description of how to resolve patient permission and confidentiality of clinical data

#### Reference

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.

#### Questions?