Employing Data Science to Improve Children's Health

AI, Machine Learning, and Advanced Statistical Modeling

Steve Rust

May 10, 2019

NATIONWIDE CHILDREN'S

When your child needs a hospital, everything matters.℠

HIMSS

CENTRAL & SOUTHERN OHIO *Chapter*

Topics for Today

1. Putting AI into a healthcare context
2. Quantitative methodologies for AI implementation
3. How to choose an AI method
4. Planning for implementation
5. Sustaining a Data Science program
6. Examples of AI projects in a pediatric healthcare setting
7. Other ongoing projects

Image source: https://meetingking.com/how-to-create-a-meeting-agenda//

Putting AI into a Healthcare Context

What is AI?

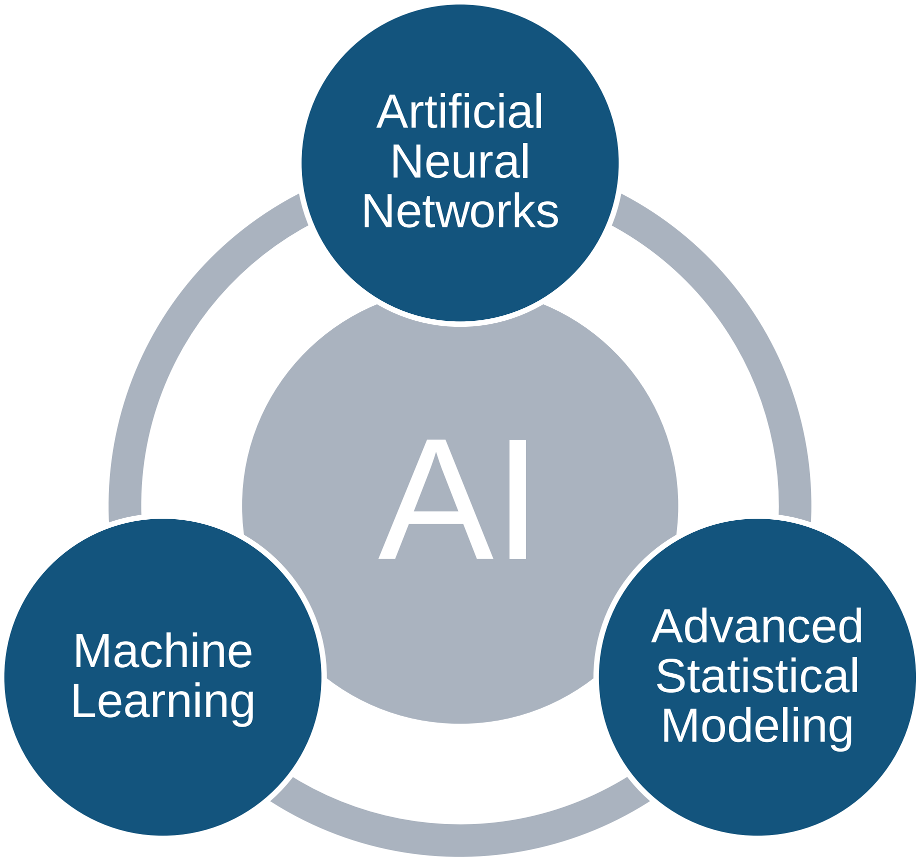
- The capability of a machine (computer) to imitate intelligent human behavior

What types of human behaviors can be effectively imitated by a computer within a healthcare setting?

- Deciding which patients in a population would benefit from an intervention program
- Anticipating adverse events for hospital inpatients
- Annotating images
- Scoring of patient assessments
- Identifying future high utilizers of healthcare services
- Retrieving relevant data from large information collections

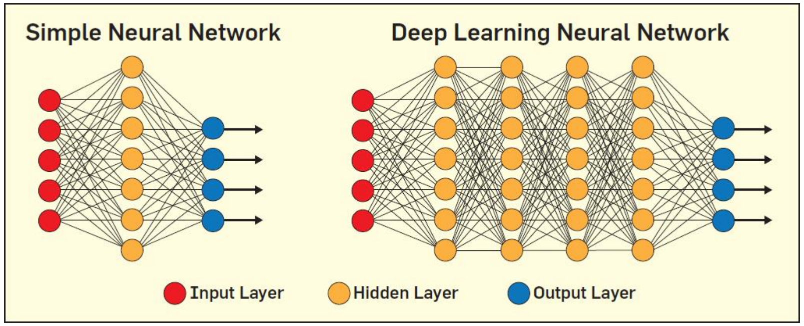
Image source: https://www.cio.com/article/3387618/ai-technology-when-to-build-when-to-buy.html

Quantitative Methodologies for AI Implementation



Artificial Neural Networks

- Large number of processing nodes arranged in layers
  - Input layer
  - Hidden layers
  - Output layer
- Trained to perform a task by being fed large amounts of labeled data



- Little or no pre-processing of data is required
- Resulting algorithms are "black box" in nature, although some progress is being made on this front
- Enabled by GPU-based hardware & software advances
- Current buzzphrase – Deep Learning

Image source: https://cacm.acm.org/magazines/2018/6/228030-deep-learning-hunts-for-signals-among-the-noise/fulltext?mobile=false

Machine Learning Methods

- Popular methods:
  - Gradiant boosting trees
  - k nearest neighbors (kNN) classifiers
  - Naïve Bayes classifiers
  - Random forests (bagged trees)
  - Support vector machines (SVM)
- Best suited for moderately large training sets
- Little or no pre-processing of data is required
- Resulting algorithms are mostly “black box” in nature
- Moderate hardware requirements
- Currently quite popular – Gradient boosting

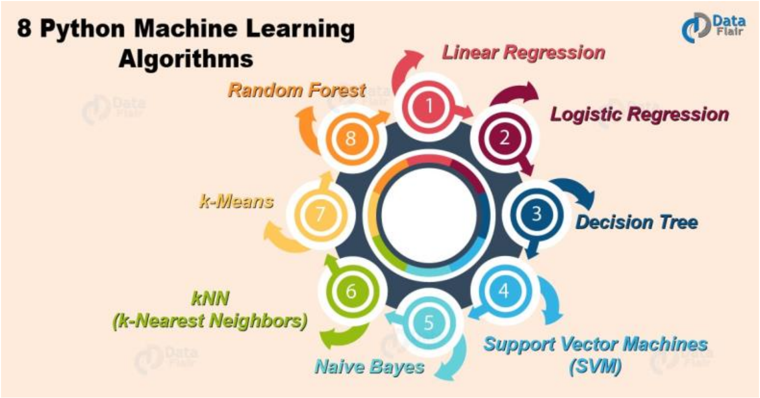
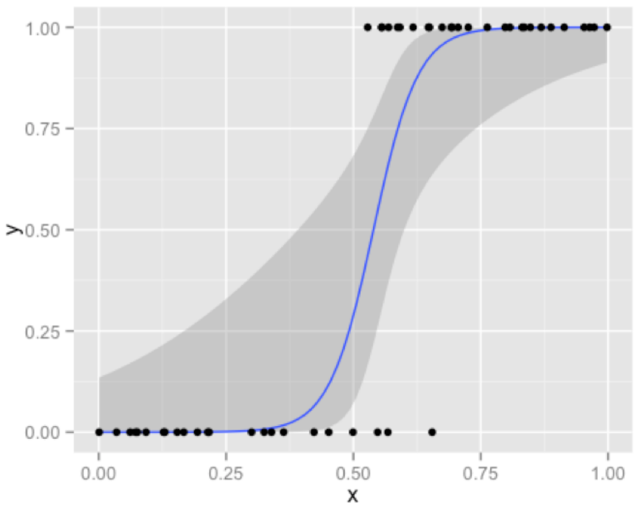


Image source: https://data-flair.training/blogs/machine-learning-algorithms-in-python/

Statistical Methods

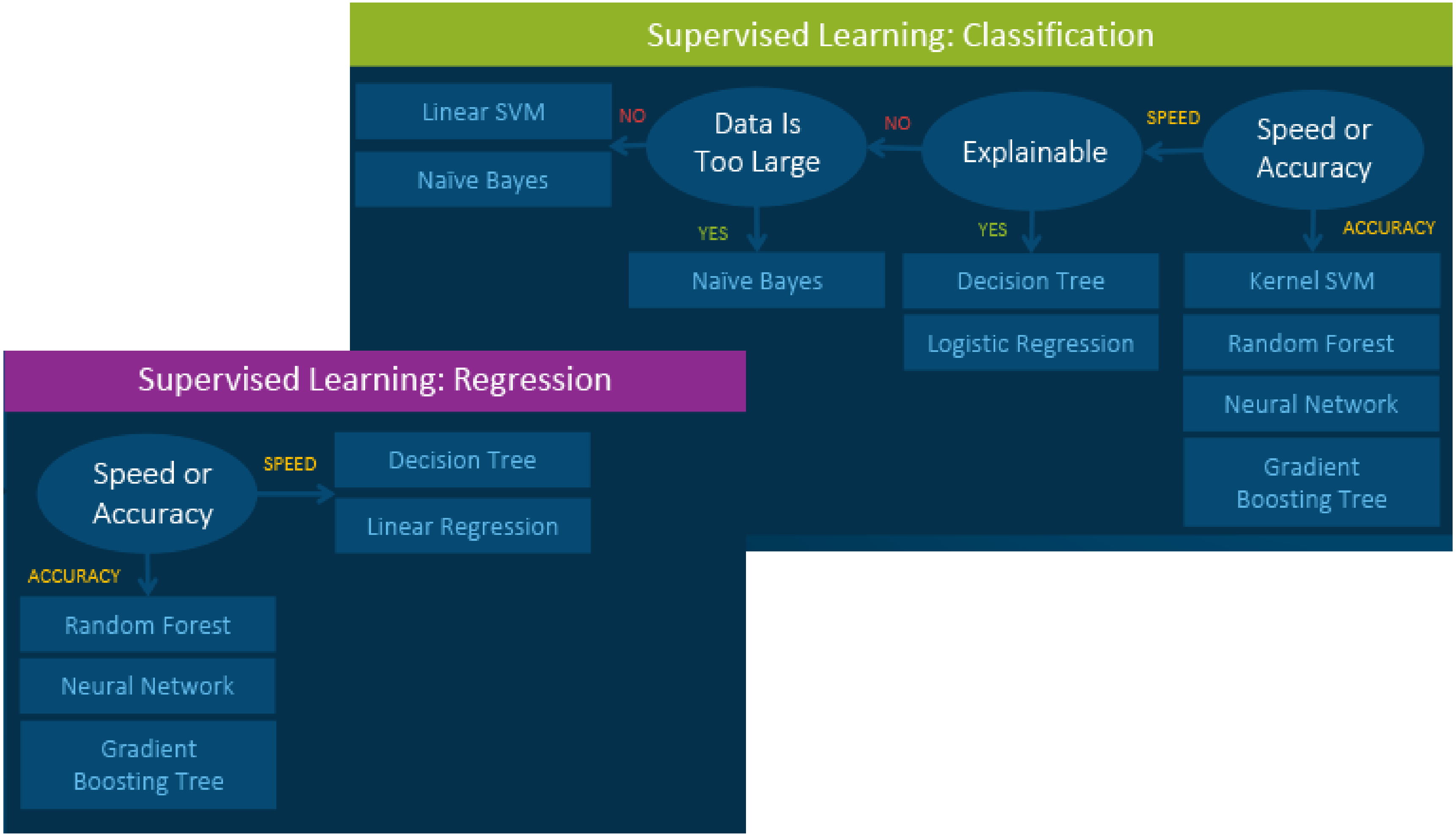
- Regression
  - Linear regression
  - Regression trees
- Classification
  - Logistic regression
  - Classification trees
- Do not require large amounts of training data



- Regularization approaches (e.g. lasso) available for regression methods
- Feature generation is required
- Resulting algorithms are “explainable”
- No hardware/software limitations

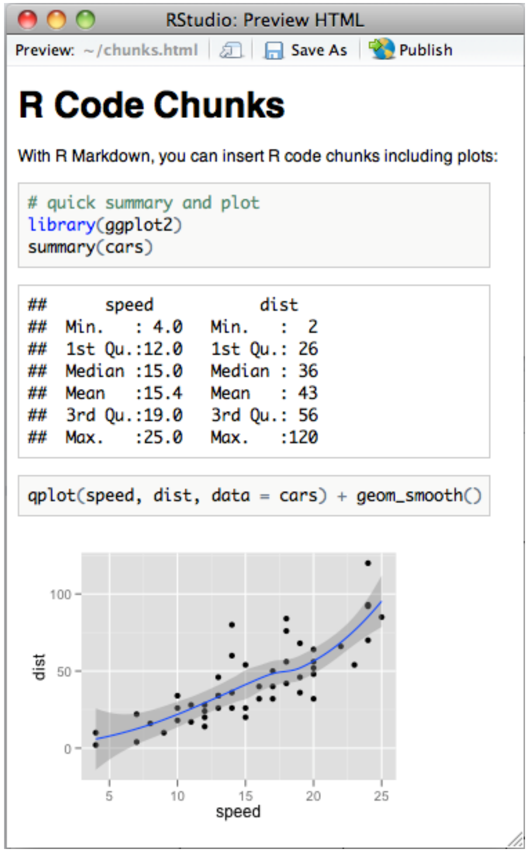
Image source: https://www.edvancer.in/logistic-regression-vs-decision-trees-vs-svm-part2/

How to Choose an AI Method



Source: https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/

Planning for Implementation



- You probably have to worry about more than speed and accuracy
- How will your algorithm be implemented so that it is integrated into the clinical or operational workflow?
- Will the algorithm be implemented within an EHR system?
  - If so, what types of algorithms can be implemented?
- Similar issues with most operational systems
- Just because you can train and evaluate the a model in R doesn't mean you'll be able to implement the resulting algorithm

Image source: https://www.rstudio.com/online-learning/

Sustaining a Data Science Program

- Form a steering committee of high-level stakeholders
- Use steering committee to prioritize use of data science resources
  - First, prioritize project concepts for 2-3 page proposal development
  - Then prioritize proposed projects for execution
- Report back to steering committee on value created in terms they will understand (e.g., better patient outcomes)
- Make internal presentations highlighting early and ongoing wins

Examples of AI Projects in a Pediatric Healthcare Setting

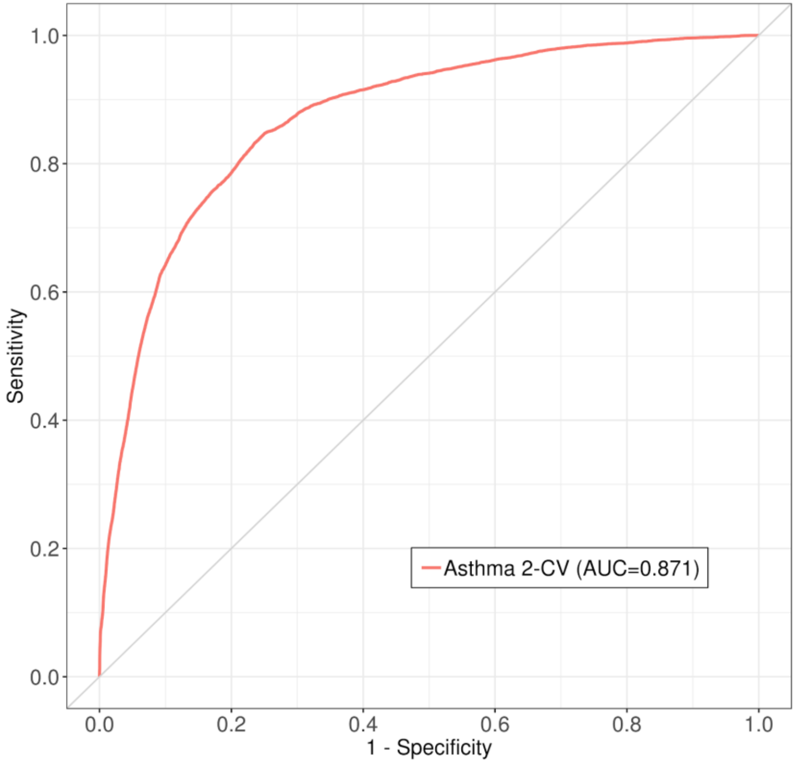
1. Predicting asthma ED visits
2. Predicting dental caries at first dental visit
3. Predicting adverse inpatient events outside the ICU
4. Annotating images to identify post-TBI activities
5. Assessing infant motor function based on MS Kinect video
6. Identifying future high utilizers of healthcare services
7. Query expansion for better clinical note searching

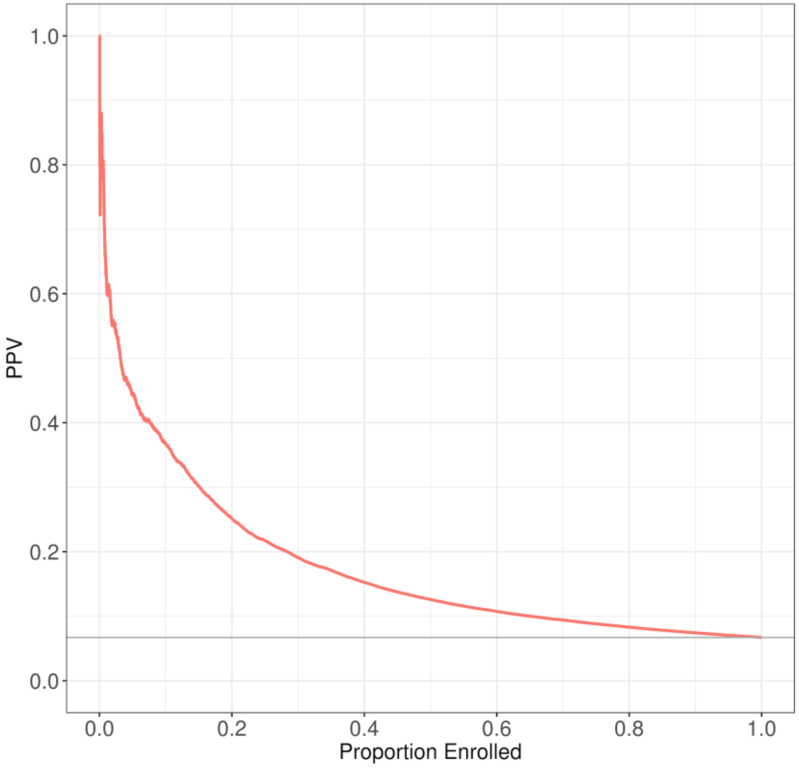
Image source: https://www.nursetheory.com/standard-nursing-gear-and-equipment-list/

Predicting Asthma ED Visits

- **Objective:** Build an algorithm that predicts the probability of an asthma ED visit within one year of a Primary Care patient encounter
- **Data:** 30 candidate predictors identified by clinical subject matter experts
- **Methods:** Fit a lasso regularized logistic regression model and characterize expected performance with 10-rep, 10-fold crossvalidation
- **Project End Goal:** Prioritize high risk patients for enrollment into school-/home-based interventions
- **Results:** 87% area under the ROC curve; among the 10% with highest risk, 37% would visit the ED for asthma within the next year

Predicting Asthma ED Visits (Cont.)

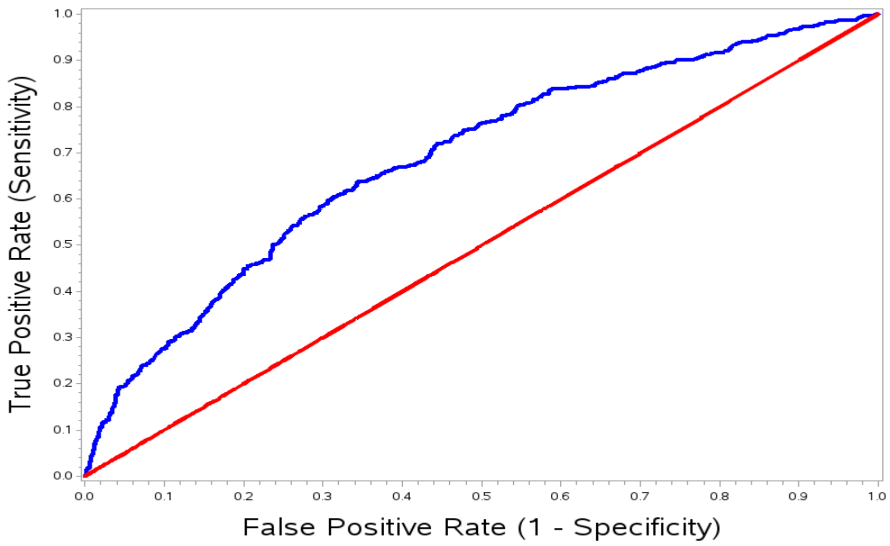


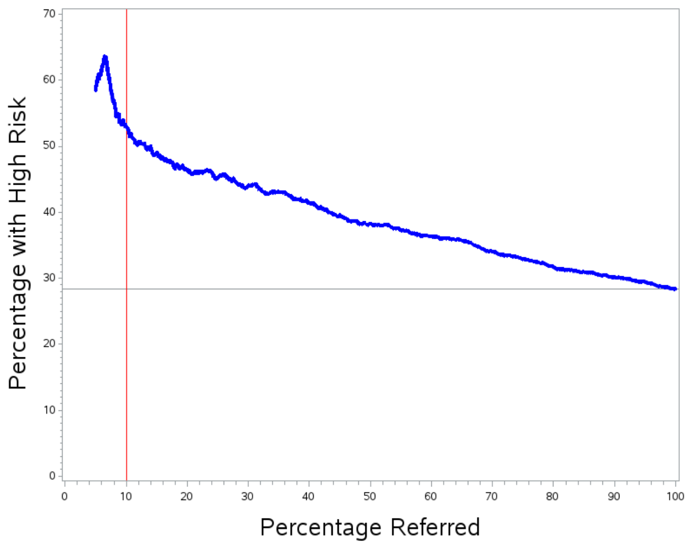


Predicting Risk of Dental Caries at First Dental Visit

- **Objective:** Develop a model that estimates the probability that a child will have dental caries at the time of their first dental visit
- **Data:** 72 candidate predictors identified by dental subject matter experts & present in the electronic medical record
- **Methods:** Stepwise (backward) logistic regression model followed by validation with 30% hold-out test data set
- **Project End Goal:** Create a simple caries risk assessment that can be assessed during primary care visits with minimal burden on the primary care physician
- **Results:** 69% area under the ROC curve; among the 10% with highest risk, 53% would have caries at their first dental visit

Predicting Risk of Dental Caries at First Dental Visit (Cont.)





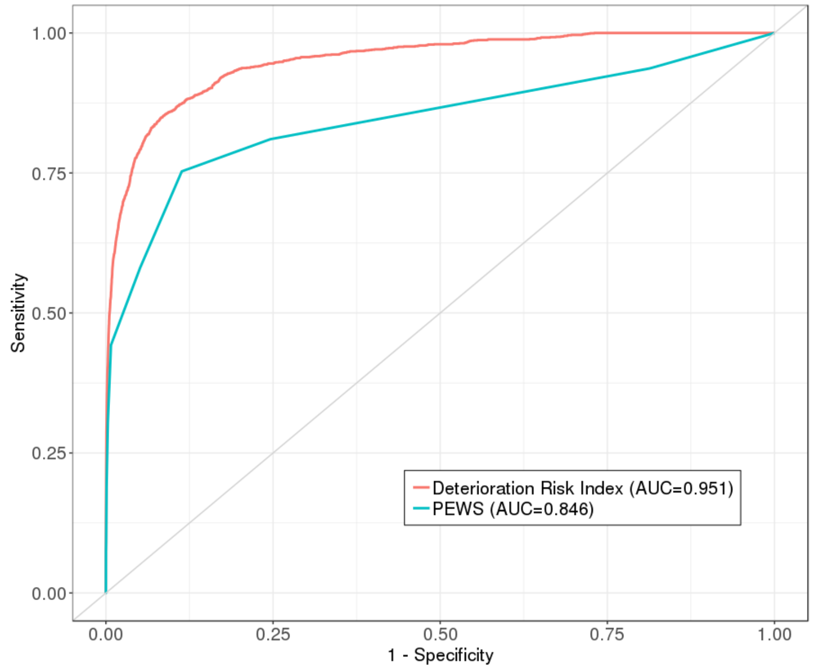
HIMSS CENTRAL & SOUTHERN OHIO Chapter

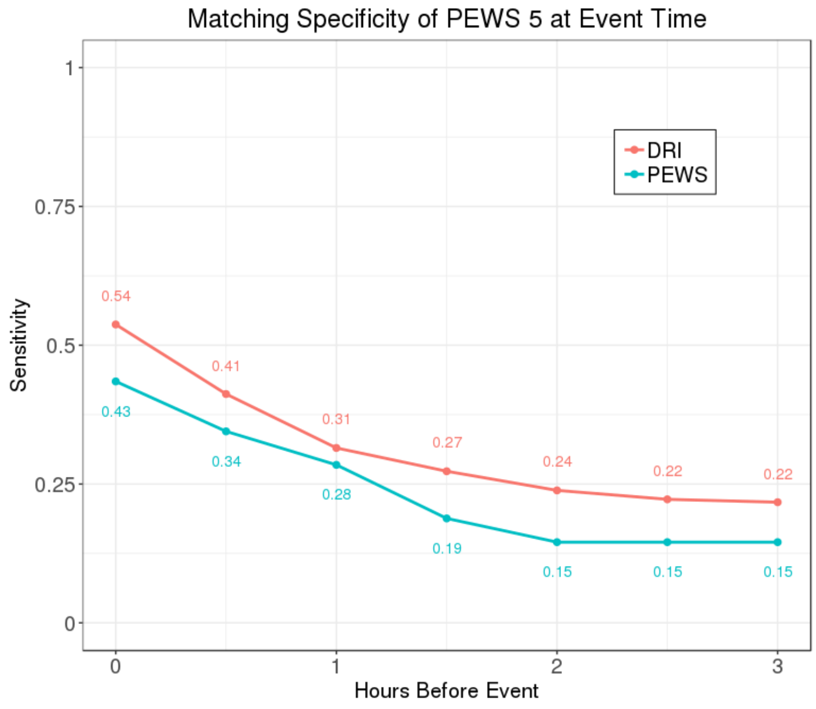
Predicting Adverse Inpatient Events Outside the ICU

- **Objective:** Develop a algorithm (DRI) that estimates the probability that an non-ICU inpatient will have a serious adverse event within the next 24 hours; target adverse events are:
  - Cardiopulmonary failure (Code Blue)
  - Emergency transfer to the ICU
  - Death
- **Data:** 85 candidate predictors identified by clinical experts
- **Methods:** Fit a lasso regularized logistic regression model and characterize expected performance with 10-rep, 10-fold crossvalidation; compare performance with current early warning system (PEWS)
- **Project End Goal:** Prioritize high risk patients for enrollment into "watcher" intervention program

Predicting Adverse Inpatient Events Outside the ICU (Cont.)

- **Results:** 95%% area under the ROC curve; ability to provide 2-3 hours advance warning allowing time for effective risk mitigation





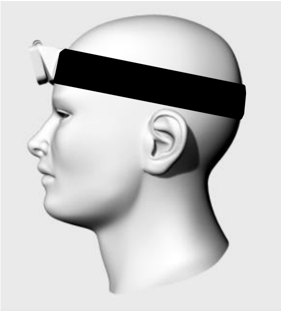
Annotating Images to Identify Post-TBI Activities (Cont.)

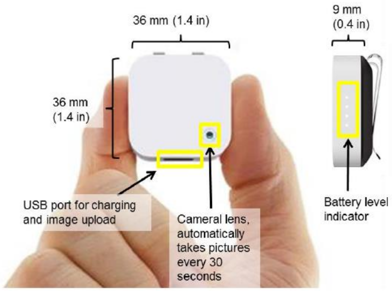
- **Overall Objective:** Collect and analyze objective, real-time data on physical and cognitive rest following sports-related concussions among youth
- **Physical Rest Data:** Surveys plus

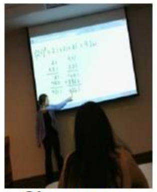

Activity monitor

Heart Rate Sensor Strap

- **Cognitive Rest Data:** Surveys plus Narrative Clip images






Classroom

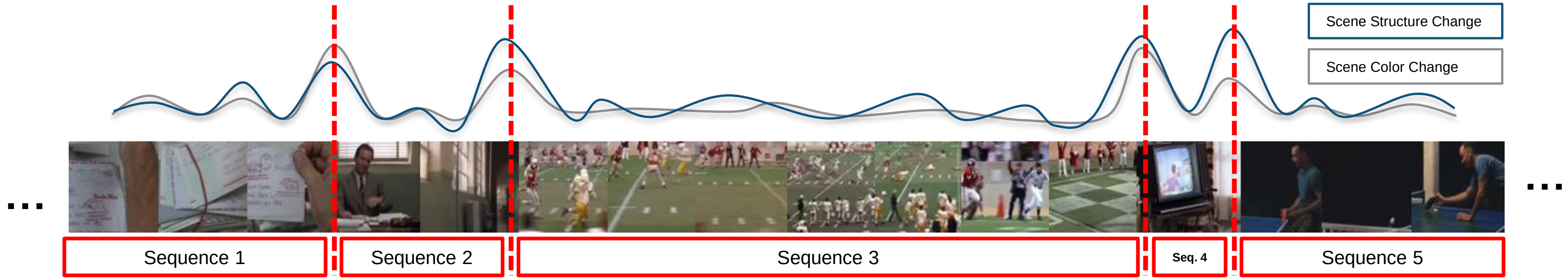

Homework


Video Gaming


Texting

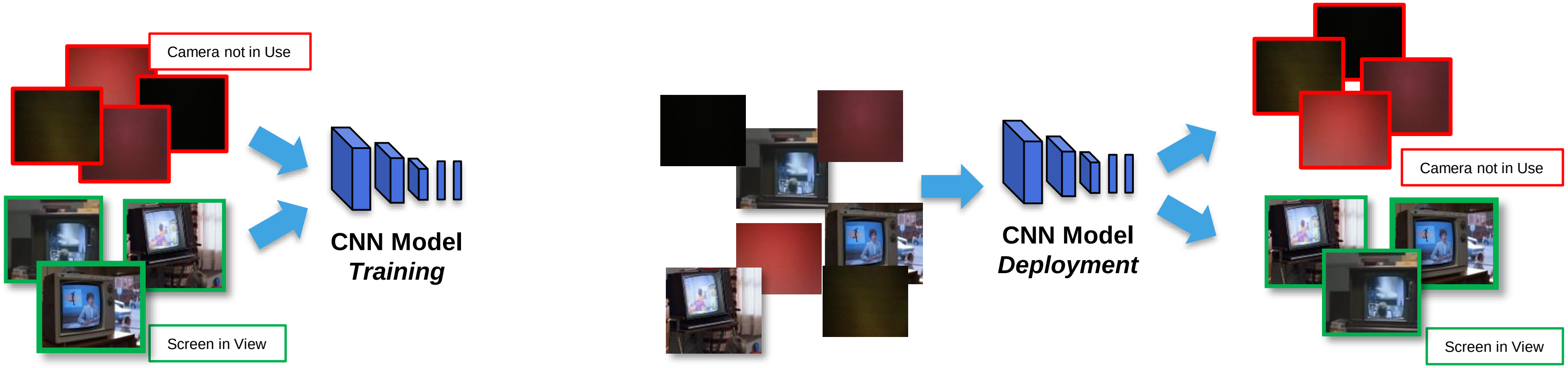
Annotating Images to Identify Post-TBI Activities

- **Problem:** Time lapse data coding of Narrative Clip image streams is very labor-intensive
- **Solution:** Employ convolutional neural networks to pre-process the image streams
  - Divide the image stream into segments of similar activity



Annotating Images to Identify Post-TBI Activities

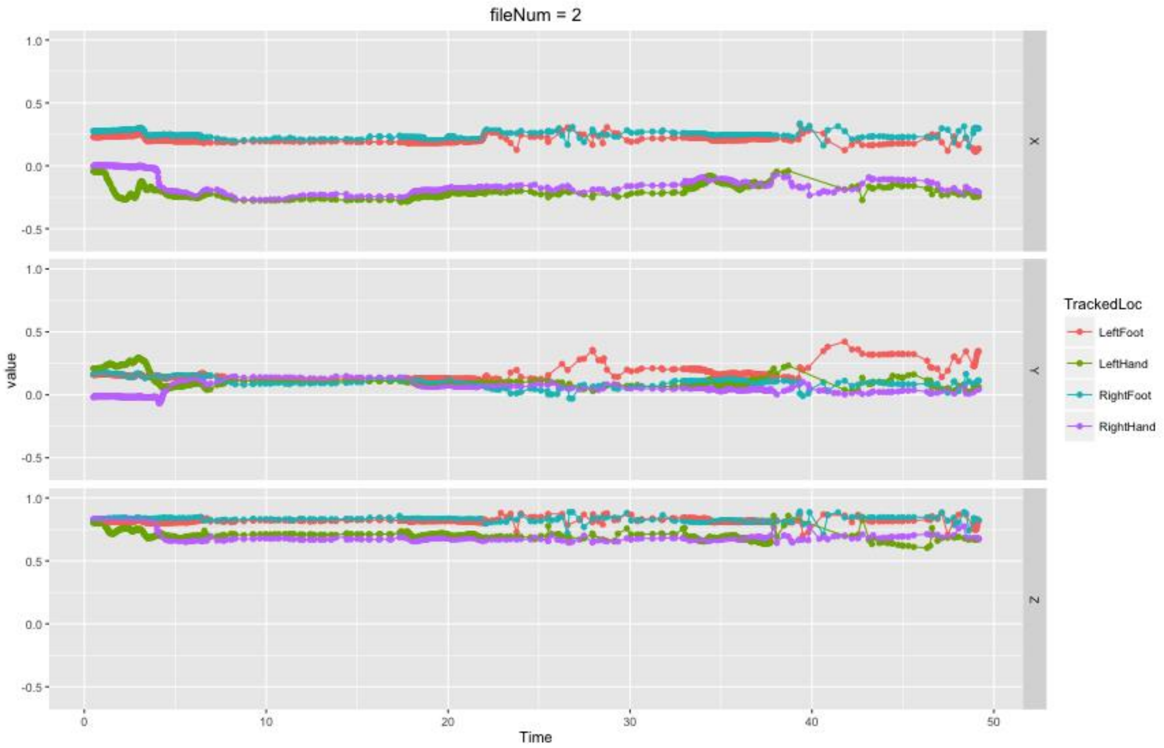
- Attach one of 3 labels to each image
  - Camera off
  - Screen in view
  - No screen in view



- **Result:** More efficient manual coding of Narrative Clip image streams

Assessing Infant Motor Function Based on MS Kinect Video

- **Data Collection:** Motion-tracking video captured by MS Kinect camera while infants are encouraged to move extremities for 2 minutes

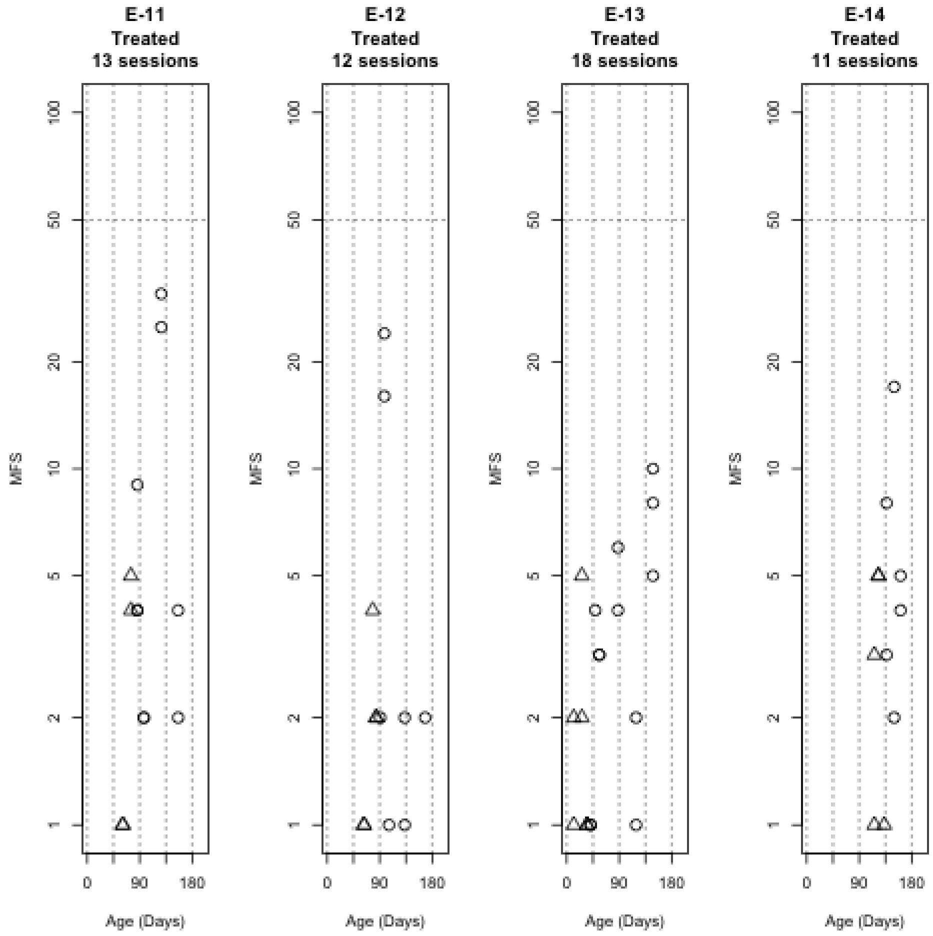


Assessing Infant Motor Function Based on MS Kinect Video

- **Objective:** Develop algorithm based on the motion-tracking data that produces a motor function score (MFS) in the range 0-100; MFS should discriminate infants with SMA from healthy infants
- **Feature Engineering**
  - Motion features were extracted for five motion feature classes (distance, direction, direction change, velocity, acceleration)
  - For each motion feature class:
    - The feature space was divided into 100 regions via cluster analysis
    - A histogram was created for each session to record the proportion of the time the data lies within each of the 100 regions
  - The result is 500 features for modeling
- **Modeling Approach:** Fit SVM that discriminates infants with SMA from healthy infants; use model probability of healthy infant as MFS

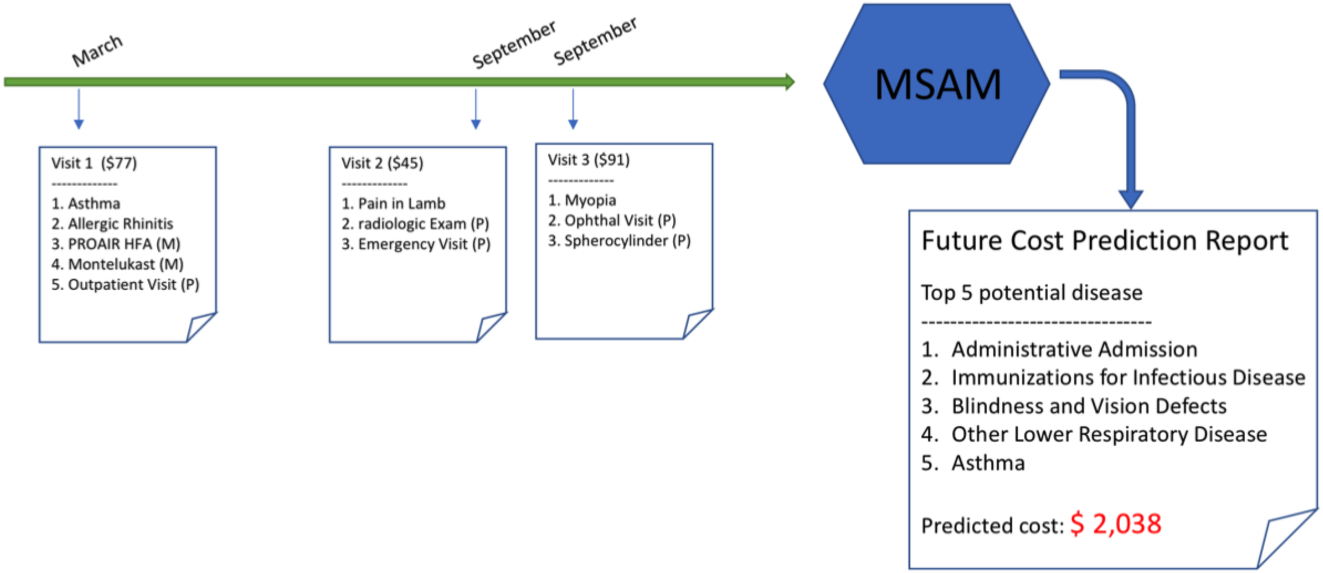
Assessing Infant Motor Function Based on MS Kinect Video (Cont.)

- **Results:** Almost perfect discrimination achieved & MFS can be used to characterize response to gene therapy treatment



Identifying Future High Utilizers of Healthcare Services

- **Objective:** Identify future high utilizers (next year) based on claims data (from this year)



Identifying Future High Utilizers of Healthcare Services (Cont.)

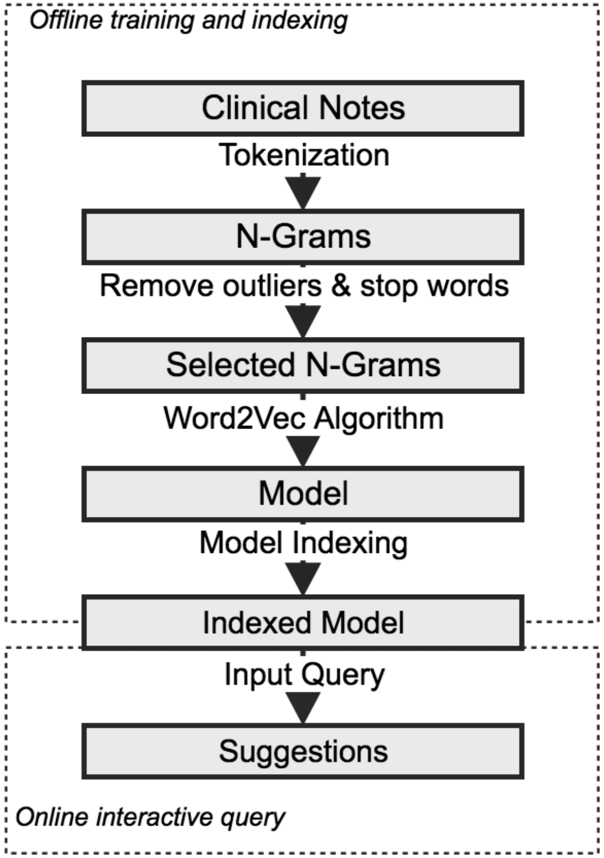
- **Approach**
  - Project medical codes into vector representation
  - Aggregate medical code vectors to form visit representation
  - Aggregate visit vectors to form patient representation
  - Use patient vector for outcome prediction
  - Employ different forms of neural network models to make predictions
- **Results**

PFK	MAE	MAE (censored@50K)	R^2	R^2(censored@50K)
Prior Year Cost	0.7612	0.6989	0.061	0.001
MLP	0.8569(0.036)	0.7669(0.031)	0.1080(0.058)	0.0937(0.059)
RNN	0.770(0.049)	0.6884(0.033)	0.1211(0.071)	0.1921(0.056)
Bi-RNN	0.743(0.038)	0.6910(0.026)	0.1397(0.065)	0.1421(0.065)
Att-Bi-RNN	0.798(0.032)	0.7110(0.039)	0.1797(0.081)	0.1984(0.045)
Timeline	0.7301(0.031)	0.6354(0.022)	0.1432(0.078)	0.1782(0.053)

Query Expansion for Better Clinical Note Searching

- **Background**
  - A large amount of clinical data is stored in notes
  - Clinical notes frequently contain spelling variations, typos, local practice-generated acronyms, synonyms, and informal words
    - E.g. “tonsillectomy” can miss notes containing “tonsilectomy” (common misspelling), “t/a” and “t&a” (nonstandard but commonly used abbreviations), or “adenotonsillectomy” (semantically related concept)
  - Ontology-driven approaches lack timely updates and only cover formal words
- **Solution – Deep Suggest**
  - We offer suggestions to expand and pivot queries to help overcome these challenges, including word variations (e.g., formal or informal forms, synonyms, abbreviations, misspellings) and other relevant words (e.g., related diagnosis, medications, and procedures)
  - Human intelligence is then used to further refine or pivot their query

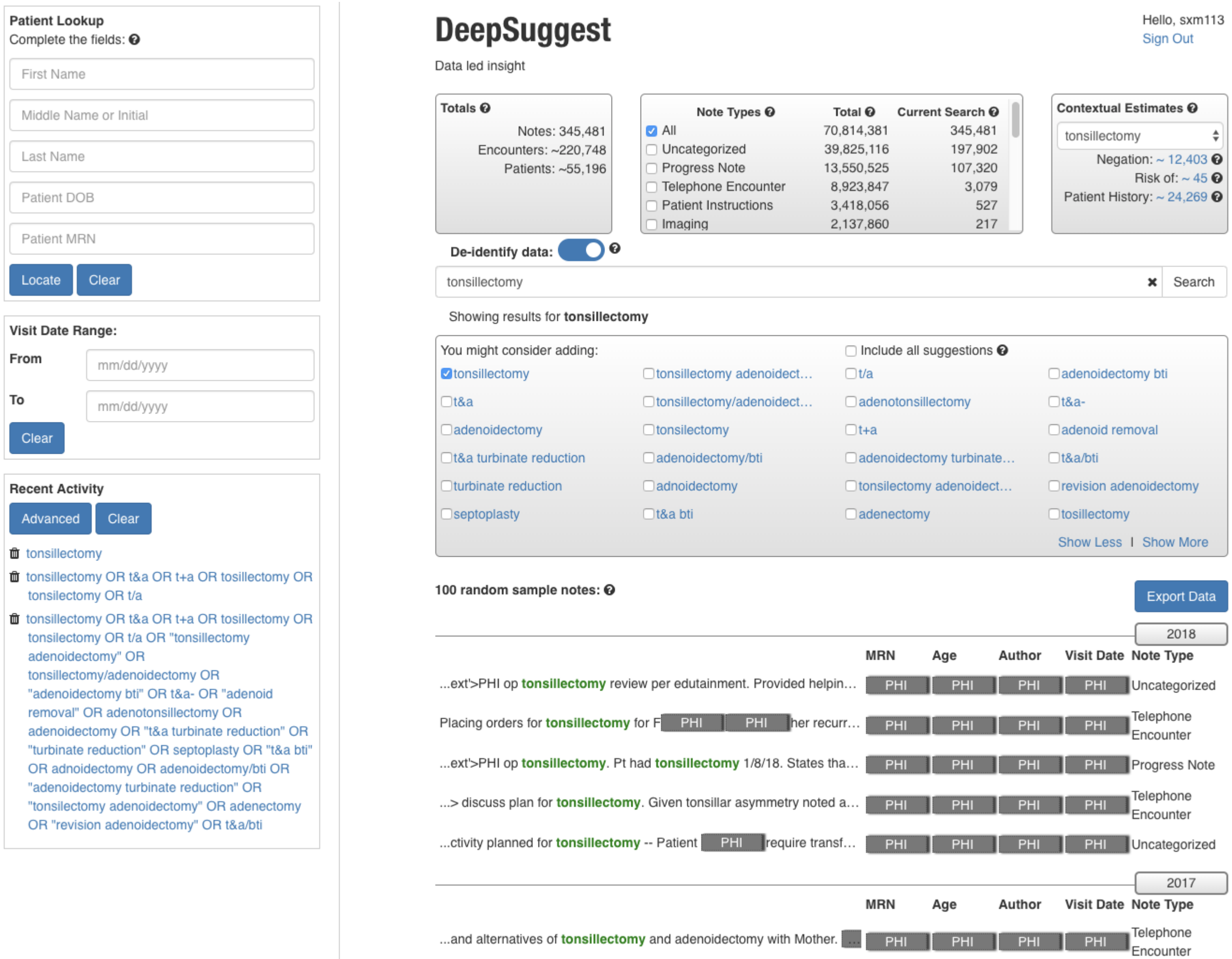
Query Expansion for Better Clinical Note Searching (Cont.)



- **Methodology**
  - Word2Vec (unsupervised neural network) on 1-4 grams
  - Corpus: 69GB of over 66 million clinical notes from 2006 until end of 2016
  - After data preprocessing and excluding words with less than 30 occurrences, our vocabulary consisted of 6.3 million unique 1-4 gram words, representing 5.5 billion total words
- **Results**
  - P@60 = 72%. Usability test resulted that DeepSuggest is able to achieve two-times recall on clinical notes compared to Epic

HIMSS CENTRAL & SOUTHERN OHIO Chapter

Query Expansion for Better Clinical Note Searching (Cont.)



HIMSS CENTRAL & SOUTHERN OHIO *Chapter*

Other Ongoing Projects

- Neonatal abstinence syndrome (NAS) scoring
- Non-accidental trauma (NAT) detection
- Necrotizing enterocolitis (NEC) prediction
- No-show prediction
- Risk of opioid use disorder
- Seizure detection based on wearable device data
- Sepsis prediction
- Suicide prediction

Image source: https://cuddlesandmuddles.wordpress.com/2014/06/10/so-much-to-do-so-little-time-to-do-it-in/

NATIONWIDE CHILDREN'S

When your child needs a hospital, everything matters.℠

Questions?