Mining Healthcare Systems to Personalize Patient Care & Improve Clinical Decisions

Knowledge Solutions
Siemens Medical Solutions USA, Inc.

R. Bharat Rao
Glenn Fung
Balaji Krishnapuram
Jinbo Bi
Murat Dundar
Vikas Raykar
Shipeng Yu
Sriram Krishnan
Xiang Zhou
Arun Krishnan
Marcos Salganicoff
Luca Bogoni
Matthias Wolf
Anna Jerebko
Jonathan Stoeckel

Copyright © 2010 Siemens Medical Solutions USA, Inc. All rights reserved.
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Mining Medical Images
  - Quality
  - Personalized Medicine
- Conclusion
Global Healthcare Market Trends

- Increasing Healthcare Costs (73% over the next 9 years)\(^1\)
- Increased consumerism & consumer awareness
- New technologies

Knowledge explosion
  - Increased creation and dissemination of evidence based guidelines

Data overload
  - Increased digitization of data

Enables Personalized Medicine

Opportunities for Data Mining

---

\(^1\) "National Health Expenditures...Calendar Years 2009—2018" - Centers for Medicare & Medicaid Services, Office of the Actuary
Clinical decisions are more complex

Clinicians need support to manage these challenges. Data mining and personalization of care can help.
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Mining Medical Images
  - Quality
  - Personalized Medicine
- Conclusion
Integrate patient data with medical knowledge to improve health outcomes.

- Access and **mine existing patient data** from disparate sources
- Integrate **learned knowledge** for decision support
- Impact workflow & **outcomes**
Impacting Clinical Workflow

Data → Knowledge Solutions → Clinical Answers

Learned Knowledge
REMIND™ Knowledge Platform*: Architecture
Reliable Extraction & Meaningful Inference from Nonstructured Data

Images
Patient Factors
Proteomics
Genomics
Treatment Plans

Extraction

REMIND Platform
Combine Conflicting Local Evidence
Probabilistic Inference Over Time

Extraction

Extraction

Extraction

Extraction

Plug-in Learned Knowledge

*Not considered a Medical device.
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Mining Medical Images
  - Quality
  - Personalized Medicine
- Conclusion
Mining Medical Images
Computer-Aided Detection (CAD)

REMIND Platform

Data → Clinical Answers

Images

Knowledge

Detected structures
- 2nd reader for physicians

Mined knowledge
- Predictive models discovered from patient data
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Mining Medical Images
  - Quality
  - Personalized Medicine
- Conclusion
Automated Quality Abstraction & Reporting

Text Mining

Manual Chart Abstraction

Clinical Notes

Lab Data

Pharma Data

REMIND™ Platform

CMS measures

Quality Measures enabled by the REMIND Platform

Copyright © 2010 Siemens Medical Solutions USA, Inc. All rights reserved.
“Soarian Quality Measures is an essential, time-saving solution that automates the abstraction of critical patient data from both structured and unstructured free text from clinical narratives. It enables our teams to focus on providing the best quality of care possible.”

Janene Yeater, Vice President of Quality and Planning MedCentral Health System - July 2009

Quality Measurement & Improvement

Data → REMIND Platform → Clinical Answers

Automated Quality Measurement & Improvement

Knowledge

Existing guidelines
- CMS & Joint Commission guidelines
Mined from free text

Labs
Demographics
Patient Factors
Medication
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Mining Medical Images
  - Quality
  - Personalized Medicine
- Conclusion
Combining clinical data from disparate sources improves prediction accuracy

LOO ROC Plot for S2y (82pts, P/N: 24/58)

- AUC: 0.65 (Clinic)
- AUC: 0.76 (Clinic + Image)
- AUC: 0.85 (Clinic + Image + Biomarker)

Personalized dose recommendation
Non Small-Cell Lung Cancer

Current plans based on “average dose response curves” appropriate for a population

REMIND Platform: Prediction models for lung cancer patients: www.predictcancer.org

Create personalized dose response rate for individual patients for models created by mining large database of patient data and external medical knowledge

*This information about this product is preliminary. The product is under development and not commercially available for sale in the U.S. and its future availability cannot be ensured.
What about data from multiple institutions?

Predicting 2 year Lung Cancer Survival

- 455 inoperable NSCLC patients, stage I-IIIB, referred to the MAASTRO clinic (Netherlands) to be treated with (chemo)radiotherapy
- 112 patients from Gent hospital (Belgium)
- 40 patients from Leuven hospital (Belgium)

How do we use data from different institutions while preserving patient privacy?
Example: Cox Regression for Survival Analysis

- Cox regression, or the Cox propositional-hazards model, is one of the most popular algorithms for survival analysis
  - Defines a semi-parametric model to directly relate the predictive variables with the real outcome (i.e. the survival time)
  - Assumes a linear model for the log-hazard:
    \[ \lambda(t|\mathbf{x}_i) = \lambda_0(t) \exp[\mathbf{w}^T \mathbf{x}_i] \]

- We propose privacy-preserving Cox regression (PPCox) which is based on random projection
  - Provides accurate classification
  - Does not reveal private information
    \[ \lambda_{\text{HPCCox}}(t|\mathbf{x}_i) = \lambda_0(t) \exp[\mathbf{w}^T \mathbf{B}^T \mathbf{x}_i] \]
We are using real data from several institutions across Europe to build models for survival prediction for non-small-cell lung cancer patients while addressing the potential privacy preserving issues.

Example showing the improvement in performance of a model trained using all the data available from multiple sites against models learned only using local available data.
Prediction of survival outcomes and side effects

**Patient clinical factors**
- Age: 66
- Gender: female
- Height (cm): 179.00
- Weight (kg): 72.00
- BMI: 22.47
- Weight loss (during the last 3 months): stable
- Smoking: current smoker
- Pack Yrs:
- WHO Perf Scale: 1.00
- Charlson Comorbidity Index: 0.00
- Lung function FEV1, 1 second (lit): 
- Lung function FEV1, 1 second (%):
- Hemoglobin level (mmol/l): 7.70

**Tumor Information**
- Stage: IIb
- T-Stage: T4
- N-Stage: N3
- M-Stage: M0
- Pathology: 
- Histology: squamous cell ca

**Image Information**
- Gross Tumor Volume (cc): 83.78
- Maximum Standard Uptake Value, primary tumor (from FDG PET): 133.00
- Num. of pos. lymph node stations: 1.00

**Treatment Plan**
- Mean Lung Dose [Gray]: 15.82
- V20 [%]: 24.00
- Treatment regime: sequential before RT
- Overall treatment time [days]: 41.00
- Treatment dose [Gray]: 60.00
- Fraction size [Gray]: 2.00
- Fractions/day: 1.00
- Number of cycles [chemo]: 3.00
- Chemotherapy type:

**Prediction**
- Probability of Survival at 2 yrs: 34%
- Probability of pneumonitis: 11%

*This information about this product is preliminary. The product is under development and not commercially available for sale in the US and its future availability cannot be ensured.*
Increasing treatment dose increases the probability of survival, but also increases likelihood of side effects.
Further increasing treatment dose increases likelihood of side effects, but does not improve probability of survival.
Which asymptomatic patients will develop hypertension within 5 years?

- Based on data from large study, >4000 patients over 10 years
  - Genetic (SNPs), clinical, lab, imaging, ...
- Analyzed 100s of potential predictive variables to identify the most informative subset
- Learned a Bayesian Network
  - predicts patients who are at high risk to become hypertensive (AUC >0.85).
- Model also suggests patient specific interventions that can reduce risk for a subset of patients,
- Undergoing multi-site validation

Targeted Decision Support at Point of Care
Personalized therapy selection for Lung cancer

REMIND Platform

Data → Clinical Answers

Knowledge
- Treatment impact
  - Adjust dose based on outcomes & side effects

Existing guidelines

Learned knowledge
- Predictive models discovered from patient data

Labs Images

Treatment Plans Genomics

Copyright © 2010 Siemens Medical Solutions USA, Inc. All rights reserved.
Creating Medical Knowledge
Mining large patient databases from multiple institutions

Data -> REMIND Platform -> Predictive Models

Knowledge

Existing guidelines for cancer therapy

Learned knowledge
- Predictive models discovered from patient data
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Mining Medical Images
  - Quality
  - Personalized Medicine
- Conclusion
The Road to Personalized Medicine

Knowledge-Driven Workflow
Clinical decisions using all available electronic data & medical knowledge for Personalized Medicine, Quality Improvements, and Workflow Management.

Key Enablers
Learn new knowledge from patient records
Exploit existing knowledge

Focused Tasks
"Does this patient have a history of smoking?"
"Find a lung nodule in a chest CT"