UNIVERSITY of MARYLAND
SCHOOL OF MEDICINE

Clinical Applications of Big Data

Michael A. Grasso, MD, PhD, FACP

Assistant Professor
Internal Medicine, Emergency Medicine, Computer Science
University of Maryland School of Medicine

Director
University of Maryland Clinical Informatics Group
www.UMCIG.com

Department of Emergency Medicine
110 S. Paca Street, 6th Floor, Suite 200
Baltimore, MD 21201
mgras001@umaryland.edu

Outline

- Big Data
  - Clinical Decision-Making
  - Big Data Challenges
  - Sources of Big Data
  - Our Approach
  - Areas of Research
- Projects
  - Knowledge Representation and Reasoning
  - Patient Safety in Emergency Medicine
  - The Nature of Clinical Expertise
  - Pre-Hospital Syndromic Surveillance
  - Chronic Disease Prediction with Genomic Data
  - Computational Image Classification

The Challenge

- Systems to enhance practice of medicine.
  - Physician-driven clinical challenges.
  - Deliver safer and more efficient care.
  - Enable decision support at the bedside.
- Strategic importance to the UMMC and UMSOM.
  - Enhance access to biomedical knowledge.
  - Strong theoretical basis in Computer Science.

Big Data - Clinical Decision-Making

- The practice of medicine.
  - "Medical practice" is "medical decision-making".
  - This is the defining skill of all physicians.
- "Diagnostic gap" in computations systems.
  - Many computational advances in healthcare.
    - Administrative, workflow, imaging, devices, etc.
  - Few advances in bedside clinical decision support.
    - Some success with alerts, calculators, and order sets.
    - But no computationally-enabled clinical decision support.

There are no practical systems to help doctors make clinical decisions.

Big Data - Challenges

- Accumulating data faster than we can analyze.
  - Clinicians require immediate access to 2-5 million facts.
  - Medical knowledge doubling every 5 years.
  - Clinical data doubling every 1-2 years.
- Analytical challenges.
  - Dimensionality, heterogeneity, interdependency, complexity.
  - Uncertainty, nonmonotonic, nondeterministic.
- Traditional statistical approaches to big data.
  - Efficiency and accuracy problems.
  - A priori models limit ability to find hidden patterns.

Big Data - Sources

- Department of Veterans Affairs repository (VINCI).
  - 15 years of clinical data from 150 hospitals and 800 clinics.
  - 20 million patients, 6 million currently active.
- Million Veteran Program (MVP).
  - Genomic sequences and markers, correlated with VINCI.
- Electronic Maryland EMS Data System (eMEDS).
  - Assessments, treatments, and dispositions for 400,000 priority medical EMS calls annually.
- GENEVA Consortium.
  - Secondary analysis of clinical and demographic data with high-dimensional genomic markers.

Big Data - Approach

- Semantic analysis.
  - Provide context and meaning to the clinical data.
- Machine learning.
  - Reduce intractable amounts of clinical data into a moderately-sized repository of medical facts.
- Pathophysiology.
  - Organize clinical knowledge according to physiologic relationships and evidence-based guidelines.
- Human factors.
  - Incorporate an understanding on the nature of clinical expertise in decision making.

Big Data - Areas of Research

- Dimensionality reduction.
- Biological enrichment (domain information).
- Discovery of relationships with genomic data.
- Knowledge extraction from unstructured text.
- Validation approaches.
- Rare event discovery.

Research Projects

- Knowledge Representation and Reasoning (KRR)
  - Disease, critical event, and treatment efficacy prediction.
- Patient Safety in Emergency Medicine
  - Identify patient safety indicators in emergency medicine.
- The Nature of Clinical Expertise
  - Elucidate the clinical decision-making process.
- Pre-Hospital Syndromic Surveillance
  - Risk analysis for obscure syndromes and toxidromes.
- Chronic Disease Prediction with Genomic Data
  - Genomic prediction models in pre-symptomatic individuals.
- Computational Image Classification
  - Cellular communications and surgical safety.

Research Projects - KRR

- Restructure data for bedside decision support.
  - Disease prediction.
  - Critical event prediction.
  - Treatment efficacy prediction.
- Focus on a small group of chronic diseases.
  - CAD, DM, CKD, COPD, AD, Prostate + Pancreatic CA.
  - Complex and multifactorial.
  - Leading causes of morbidity and mortality.
- Strategic collaborations.
- New computing facilities at the Baltimore VA.

Research Projects - KRR Clinical Narratives

- Semantic framework for clinical decision support.
  - Apply text analytics to clinical narratives.
  - I2B2 data set (www.i2b2.org).
  - Establish relationships between extracted terms using domain-specific medical ontologies.
  - Infer additional facts using OWL reasoner with clinical rules.
- Initial results.
  - Extract evidence-based risk scores from clinical narratives.
    - TIMI Risk Score for Acute Coronary Syndrome.
    - San Francisco Syncope Rule.
  - Great than 90% accuracy.

Research Projects - Patient Safety

- Patient safety is an essential health care challenge.
  - Reporting, analysis, and prevention of medical errors.
- Safety challenges in emergency medicine.
  - 100 million annual visits.
  - High-workload, time-sensitive, nondeterministic, information-poor, disconnected, life-critical care.
- Safety events difficult to measure.
  - Events resulting in harm just the “tip of the iceberg”.
  - Need to identify “submerged” events.
    - Near misses and events that did not result in harm.
    - Hard to find with self-reporting & with a priori models.

Research Projects - Clinical Expertise

- The nature of clinical expertise in decision making.
  - Information requirements (what, when, why).
  - Clinical guidelines, clinical prediction rules, online resources.
  - Effect of time pressure and patient acuity.
  - Impact of workflow and social interactions.
  - Inter-operator variability.
- Elucidate the clinical decision-making process.
  - Observational studies, simulations, and surveys.
- Use results to help with decision-support systems.
  - Empathic and user-driven approach to development.
  - Vetting and credentialing of decision-support systems.

"People Learning" --- not just "Machine Learning"

Research Projects - Genomic Prediction

- Chronic disease prediction with genomic markers.
  - Leading causes of M&M.
  - Obscure patterns of inheritance.
  - Prediction in presymptomatic individuals = early intervention.

- Initial results.
  - Cluster models to combine relevant clinical and genomic features.
  - New genotype score comparable to clinical risk scores.
  - Demonstrated improvements in risk prediction using domain knowledge and feature selection.
  - Identified new genomic relationships using collaborative filtering and cosine similarity.

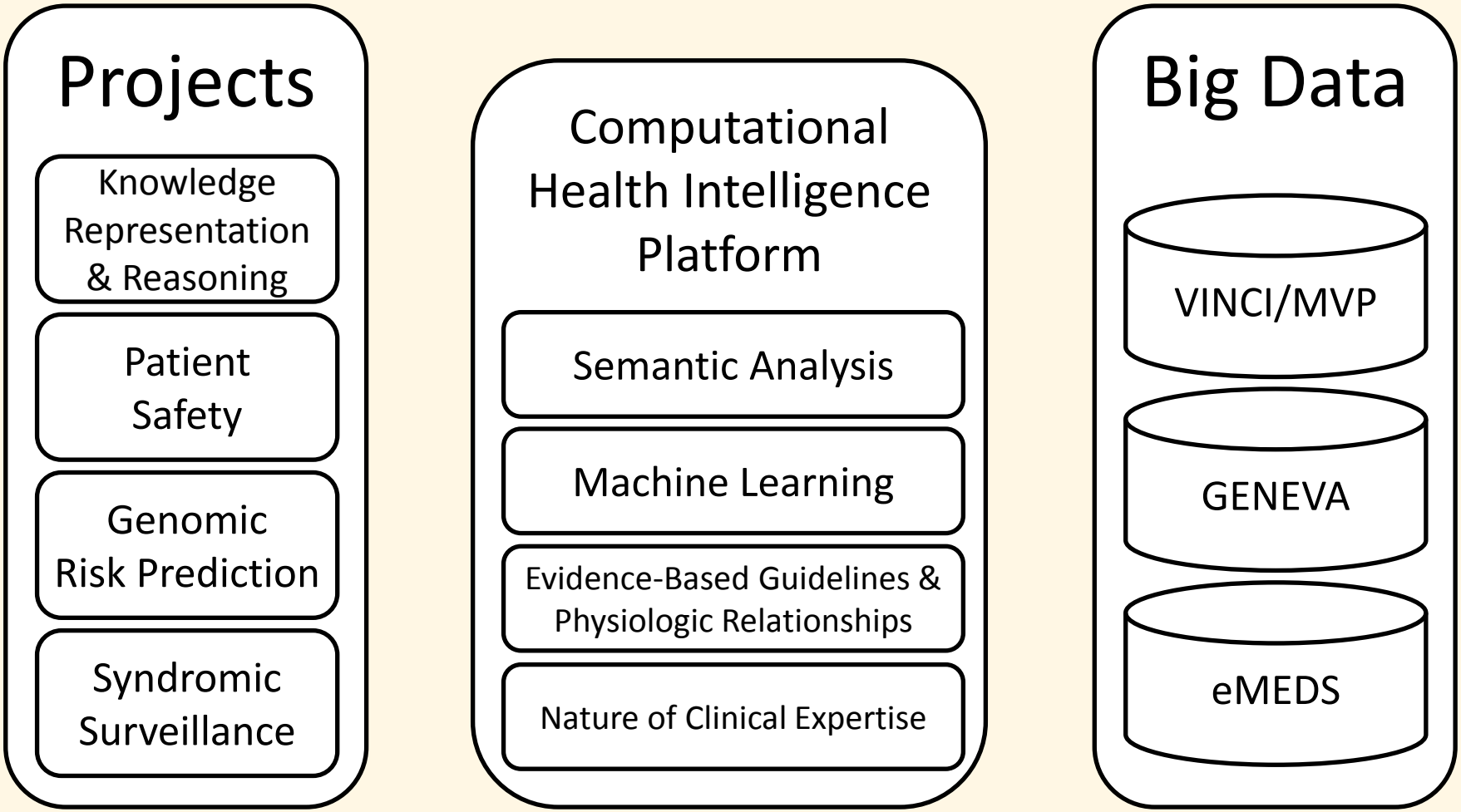
Research Projects - Syndromic Surveillance

- Early detection of disease outbreaks.
  - Biologic terrorism, disasters, or natural causes.
- Monitoring of pre-clinical data.
  - Electronic Maryland EMS Data System (eMEDS).
  - Complement with data from social media.
- Machine learning approach.
  - Identify obscure syndromes and toxidromes.
  - Predict hospital utilization requirements.

Computational Image Classification

- We experimented with image classifier techniques using machine learning algorithms.
- We developed a new approach to extract and map image features to biological characteristics.
  - Extracted image features from smooth muscle images.
  - Characterized cell-to-matrix interactions.
- We also applied this approach algorithm to laparoscopic surgery videos.
  - Identified critical surgical activities.
  - Recognized potentially unsafe actions.

Research Overview



UNIVERSITY of MARYLAND
SCHOOL OF MEDICINE

Publications

1. Grasso MA, Dalvi D, Das S, Bochare A, Kugaonkar R. Chronic disease prediction using machine learning models with genomic data. AMIA Annu Symp Proc. 2013, under review.
2. Dhariwal D, Joshi A, Grasso MA. Text and ontology driven clinical decision support system. AMIA Annu Symp Proc. 2013, under review.
3. Korolev V, Joshi A, Yesha Y, Grasso MA. On Use of Machine Learning Techniques and Genotypes for Prediction of Chronic Diseases. AMIA Annu Symp Proc. 2013, under review.
4. Bochare A, Gangopadhyay, Yesha Ye, Joshi A, Yesha Ya, Grasso MA, Brady M, Rishe N. Integrating domain knowledge in supervised machine learning to assess the risk of breast cancer. International J of Medical Engineering and Informatics, 2013, in press.
5. Bochare A, Gangopadhyay, Yesha Ye, Joshi A, Yesha Ya, Grasso MA, Brady M, Rishe N. Integrating domain knowledge in supervised machine learning to assess the risk of breast cancer. Third International Conference on Global Trends in Biomedical Informatics Research, Education and Globalization, 2012.
6. Kugaonkar R, Gangopadhyay A, Yesha Ye, Joshi A, Yesha Ya, Grasso MA, Brady M, Rishe N. Finding associations among SNPs for prostate cancer using collaborative filtering. ACM Sixth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO). 2012 Oct;: 57-60.
7. Lahane A, Yesha Y, Grasso MA, Joshi A, Park A, Lo J. Detection of unsafe actions in laparoscopic cholecystectomy videos. ACM SIGHIT International Health Informatics Symposium (ACM IHI), 2012 Jan;:315-322.
8. Martineau J, Mokashi R, Chapman D, Grasso MA, Brady M, Yesha Y, Yesha Y, Cardone C, Dima A. Sub-cellular feature detection and automated extraction of collocated actin/myosin regions. ACM SIGHIT International Health Informatics Symposium (ACM IHI), 2012 Jan;:399-407.
9. Grasso MA, Dalvi D, Das S, Gately M, Korolev V, Yesha Y. Genetic information for chronic disease prediction. IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM). 2011 Nov;:997.
10. Grasso MA. Multi-Channel Image Browser for Feature Analysis of Smooth Muscle Cells. AMIA Annu Symp Proc, 2011.
11. Lahane A, Joshi A, Finin T, Yesha Y, Grasso M. Situation-aware system for a smart operating room. IBM CASCON, 2010.
12. Grasso MA, Mokashi R, Dalvi D, Dima AA, Cardone A, Bhadriraju K, Plant AL, Brady M, Yesha Y, Yesha Y. Image classification of vascular smooth muscle cells. ACM IHI, 2010.
13. Grasso MA, Mokashi R, Dima AA, Cardone A, Bhadriraju K, Plant AL, Brady M, Yesha Y. Image classification in cell biology. AMIA Annu Symp Proc, 2010.
14. Grasso MA, Dalvi D, Dima AA, Cardone A, Bhadriraju K, Plant AL, Brady M, Yesha Y. An elliptical fit algorithm to classify the actin cytoskeleton. AMIA Annu Symp Proc, 2010.
15. Grasso CT, Finin T, Grasso MA, Joshi A, Yesha Y. A framework for the use of autonomous intelligent agents in disaster management and response. AMIA Annu Symp Proc, 2010.
16. Cardone A, Bhadriraju K, Grasso MA, Gilsinn DE, Molek M, Mokashi R, Chalfoun J, Dima AA, Brady M, Plant AL, Yesha Y, Yesha Y. Multi-channel subcellular feature analysis and correlation. Bioimage Informatics Conference September 17-19, 2010.
17. Grasso MA, Finin TW, Zhu X, Joshi A, Yesha Y. Video summarization of laparoscopic cholecystectomies. AMIA Annu Symp Proc, 2009, Nov 14-18.
18. Grasso CT, Finin TW, Grasso MA, Yesha Y, Joshi A. Intelligent Agents and UICDS. U.S. Department of Homeland Security Workshop on Emergency Management (EMWS09), Nov 5-6, 2009.