Diagnosis and Longitudinal Treatment Response

Phillip D. Levy, MD, MPH, FACEP, FAHA, FACC
Professor & Associate Chair for Research
Department of Emergency Medicine
Assistant Vice President for Translational Science and Clinical Research Innovation
Wayne State University
Disclosures

• I have the following relevant financial relationships to disclose:
  – Grant Support: NIMHD (1R01 MD005849); NHLBI (1R01 HL127215); PCORI (FC14-1409-21656)
What Problems Are We Trying to Solve?

![Diagram](image)
1. **ABILIFY** (aripiprazole)
   Schizophrenia

2. **NEXIUM** (esomeprazole)
   Heartburn

3. **HUMIRA** (adalimumab)
   Arthritis

4. **CRESTOR** (rosuvastatin)
   High cholesterol

5. **CYMBALTA** (duloxetine)
   Depression

6. **ADVAIL DISKUS** (fluticasone propionate)
   Asthma

7. **ENBREL** (etanercept)
   Psoriasis

8. **REMICADE** (infliximab)
   Crohn’s disease

9. **COPAXONE** (glatiramer acetate)
   Multiple sclerosis

10. **NEULASTA** (pegfilgrastim)
    Neutropenia
One-size fits-all medicine

Stratified medicine

Precision medicine

Stratification:
- Patients are grouped by: Disease Subtypes Demographics Clinical features Biomarkers

Personalisation:
- Patient individual: Preferences, Clinical features Medication history Environment Behaviours & habits Biomarker

Precision medicine
Exploring the Potential of Predictive Analytics and Big Data in Emergency Care

Alexander T. Janke, BS*; Daniel L. Overbeek, MD; Keith E. Kocher, MD, MPH; Phillip D. Levy, MD, MPH
Role for Personalized and Precision Medicine Tools

- Quantify baseline risk
- Monitor progression
- Refine risk prediction
- Identify disease phenotype
- Define disease mechanisms
- Personalize therapy
- Monitor response

Health Enhancement  Primary Prevention  Disease Management

Personalized Health Planning
# Limitations and Opportunities

<table>
<thead>
<tr>
<th>Limitation</th>
<th>Opportunities to Overcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parsimonious criteria</td>
<td>EMRs could provide a huge number of potential variables for deriving predictions. Data entry forms and model estimation could be designed to accept more flexible criteria.</td>
</tr>
<tr>
<td>Easy computability</td>
<td>With enhanced computational power available to clinicians, predictive models could use sophisticated techniques, such as machine learning, to compute risk estimates.</td>
</tr>
<tr>
<td>Independent validation</td>
<td>Routine data collection within EMRs, strategic prospective registries, and health information exchanges could facilitate model development and cross-validation among relatively similar populations across hospital systems.</td>
</tr>
</tbody>
</table>
“Let’s solve these first. We can worry about data mining later.”
PRECISION MEDICINE:
Integrating multi-omics, clinical and real world data

“omic” Latin suffix “ome” = mass or many.

Genome
Transcriptome
Proteome
Metabolome
Microbiome
Epigenome
Exposome

Social graph
Biosensors
Imaging

Creation of topological maps of health/disease

Eric Topol, CELL, Volume 151, Issue 1, 27 March 2014, Pages 241–253
## Strategies and Approaches

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian networks</td>
<td>Relates a set of random variables by conditional dependencies</td>
</tr>
<tr>
<td>Discriminant analysis</td>
<td>Predicts which objects fall into different classes</td>
</tr>
<tr>
<td>Decision tree learning</td>
<td>Maps observations about an item to some quantitative conclusion about that item</td>
</tr>
<tr>
<td>Matching</td>
<td>Compares differences between groups of observations that otherwise share common features</td>
</tr>
<tr>
<td>Markov and Monte Carlo simulations</td>
<td>Simulates multidimensional analyses with stochastic methods to develop approximation models</td>
</tr>
</tbody>
</table>
# Strategies and Approaches

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian networks</td>
<td>Relates a set of random variables by conditional dependencies</td>
</tr>
<tr>
<td>Discriminant analysis</td>
<td>Predicts which objects fall into different classes</td>
</tr>
<tr>
<td><strong>Decision tree learning</strong></td>
<td>Maps observations about an item to some quantitative conclusion about that item</td>
</tr>
<tr>
<td>Matching</td>
<td>Compares differences between groups of observations that otherwise share common features</td>
</tr>
<tr>
<td>Markov and Monte Carlo simulations</td>
<td>Simulates multidimensional analyses with stochastic methods to develop approximation models</td>
</tr>
</tbody>
</table>
**INTRODUCTION**
- Traditional medicine: “one-size-fits-all” approach
- Precision medicine: tailored treatment
- Challenges in tailored treatment:
  - Patient subgroup identification
  - Personalized risk factors prioritization
- Subgroup analysis is an effective approach to patient subgroup identification
  - Designed primarily for comparison of two different treatments
- Our goal: identify subgroups associated with a specific feature as risk factor
  - Feature selected according to prior knowledge

**OUR APPROACH**
- Subgroup detection tree (SDT) for detecting patient subgroups with a specific risk factor
- Motivation behind tree method:
  - Data-driven and nonparametric
  - Generate subgroups without need of prior assumption
  - Easiness in model interpretation

**ALGORITHM**
Subgroup detection tree is built in three steps:
1. Grow a large initial tree according to the splitting criteria:
   - Encourage homogeneity both in the response and the risk factor.
2. Prune the grown initial tree to obtain a nested sequence of sub-trees.
   - Seek balance between “goodness-of-fit” for training data and model complexity.
3. Select the optimal subtree:
   - Subtree with the best prediction performance on the response.

**EXPERIMENT**
- Dataset: 153 samples, 39 features
- Response: left ventricular mass indexed to body surface area (LVMi)
- Interested risk factor: vitamin D
  - Vitamin D deficiency predispose to hypertension development in African-Americans.
- Goal: detect if there exist subgroups showing strong correlation with LVMi and vitamin D

**SUBGROUP DETECTION**
- In SDT:
  - Subgroup A and D show relatively large negative correlations -0.30 and -0.40,
    indicating increasing vitamin D levels may decrease LVMi.
  - Subgroup C, LVMi is positively correlated with vitamin D, suggesting a threshold effect.
- In CART:
  - Subgroup T is the same with D, having a correlation of -0.40.
  - Other subgroups have no strong correlations.

**PREDICTION PERFORMANCE**
- Left: SDT results in four subgroups A, B, C, and D, consisting of 35, 63, 10 and 20 samples respectively.
- Right: CART generates three subgroups R, S and T.

**REFERENCE**

*Email: dzh@wayne.edu*
SDT versus RT

- **Subgroup Detection Tree**
  - COR < 252600
    - ALD < 2.65
      - Yes: A: 35
      - No: REN < 3.95
        - Yes: B: 63
        - No: C: 10
    - No: D: 20

- **Standard Regression Tree**
  - COR < 252558
    - TRIG < 193
      - Yes: R: 65
      - No: T: 20
    - No: S: 16

**Dual targets**

**Single target**
• Standard regression tree detects subgroups with guidance of a target variable

• Standard bilclustering selects subgroups without guidance of a target variable
INTRODUCTION

Personalized medicine is a multi-disciplinary area that combines data science tools and statistics techniques with medical knowledge to develop tailor-made treatment, prevention and intervention plans for individual patients.

In this study, we propose two novel data-driven approaches for risk factors selection and subgroups detection. First approach is developed based on Deep learning and second one is a new Bi-clustering method.

We applied both method in specific precision medicine problem, which focuses on subgroup of African-Americans with hypertension and poor blood pressure control who have high risk of Cardiovascular disease.

Among many features (more than 700) including demographic characteristics, previous medical history, patient medical condition, laboratory test results, and CMR results to evaluate for LVMI, 172 remained after a pre-processing step.

We used left ventricular mass index as body surface area (LVMI) as an indicator of heart damage risk. The ability to reduce LVMI would lead to an improvement in hypertension-related cardiovascular outcomes.

OBJECTIVES

Based on individual clinical data with many features, our objectives are: 1) to identify and prioritize personalized features to control and predict the amount of LVMI toward decreasing the risk of heart disease and, 2) to patient subgroup analysis to find disparities among patients which can be helpful for assigning different treatment schemes.

METHOD 1) SAFS: A DEEP FEATURE SELECTION APPROACH

We propose a new deep feature selection method based on deep architecture for high dimensional datasets. Our method uses stacked Auto-Encoders (SAE) for feature representation in higher-level abstraction. The 3 consecutive steps of our approach is as follows:

METHOD 2) SUBIC: A SUPERVISED BI-CLUSTERING APPROACH

We propose a novel patient subgroup detection method, called Supervised Bi-clustering (SUBIC) using convex optimization and apply our approach to detect patient subgroups.

Bi-clustering is defined as simultaneous clustering of both rows and columns in the data matrix.

Our approach not only finds patient subgroups with guidance of a clinically relevant target variable but also identifies subgroups by using similarity among the input variables. The consecutive steps for SUBIC approach are:

IMPLEMENTATION & EVALUATION

In this precision medicine case study, we used package H2O in R and applied our approach for different deep stacked auto-encoders. We compare our method with random forest method and LASSO. We used Mean Squared Error (MSE) as a measure for regression evaluation.

According to our results obtained from SAFS approach, the top 10 significant risk factors that affect LVMI are: Troponin Levels, Waist Circumference Levels, Plasma Aldosterone, Average Weight, Average Systolic Blood Pressure, Serum eGFR, Absolute Change, Therapeutic Intensity Score, Pulse Wave Velocity, Systolic Duration and Cholesterol Levels.

REFERENCES


Deep Feature Learning

- Myriad additive individual variables – harder to interpret and leads to overfitting
- Numbers of synergistic variable groups – easier to interpret and enhance generalization
- Deep feature learning enhances both interpretation and generalization

Stacked Auto-Encoders

• Stacked Auto-Encoder with rich representation is an appropriate deep learning architecture to learn synergistic variable groups

• An auto-encoder is trained to reconstruct its own inputs by encoding and decoding processes
## Implementation and Dissemination

<table>
<thead>
<tr>
<th>Level</th>
<th>Implementation Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td>National, regional</td>
<td>Payers, including government, may be potential investors. Existing projects, such as the National Patient-Centered Clinical Research Network and the Healthcare Cost and Utilization Project, provide precedent for database infrastructure.</td>
</tr>
<tr>
<td>Institutional</td>
<td>Institutions could use existing health information technology, in combination with prospective registries, to build usable databases.</td>
</tr>
<tr>
<td>Provider</td>
<td>Ease of access to information can be a focus. Automated message alerts and data feedback forms, carefully used, could enhance implementation and uptake. Predictive analytics might function to clarify standard of care.</td>
</tr>
<tr>
<td>Patient</td>
<td>Patient-facing applications of predictive analytics will need to have benefits that can be understood and are relevant to patients. Prediction results, carefully interpreted by clinician and patient, could be one important element of shared decisionmaking.</td>
</tr>
</tbody>
</table>
All of Us

The future of health begins with you

The Precision Medicine Initiative®
American Heart Association, Amazon Launch Cloud-Based Precision Medicine Data Marketplace

by Staff Writer 11/15/2016 0 Comments