Electronic Health Records: A New Opportunity

Discrete Events: Laboratory Interventions: Medicines, Procedures

Continuous physiologic measurements

Progress notes

Imaging Data

Interventions: Medicines, Procedures
Outline

- Risk stratification for premature infants
  - Joint work with Suchi Saria, Anna Penn, Anand Rajani, and Jeff Gould

- Prognostic stratification for breast cancer
  - Joint work with Andy Beck, Matt van de Rijn, Ankur Sangoi, Samuel Leung, Robert Marinelli, Torsten Nielsen, Marc van de Vijver, Robert West
Continuous physiologic monitoring

What information is hiding here?

<table>
<thead>
<tr>
<th>Patients</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recurring signatures discovered using unsupervised learning:

- Signatures 2 and 5 appear mostly in sick infants
- Signatures 3, 9 and 10 more common in healthy infants
- “Bad” signatures have lower entropy

Neonate Risk Stratification

What?
- Using data from first few hours of life, identify infants at risk for high morbidity
- High Morbidity (HM): Infants with short-term (e.g., culture-positive sepsis, pulmonary hemorrhage) and long-term (e.g., major hemorrhage, death)

Optimize patient care
- Parental Counseling
- Managing Staffing Ratio
- Transporting to a higher-level of care
- Comparing performance across care centers

Predicting High Morbidity

SNAP and SNAPPE-II – 12 hours (Richardson et al, J. Pediatrics 2001)

CRIB - 12 hours (International Neonatal Network, Lancet 1993)


- Performs consistently better
- Dramatically outperforms Apgar, current standard of care:
  - AUC 0.91 vs 0.69

Invasive Laboratory Tests

- Invasive lab measurements do not add to Physiscore’s predictive value

Tests added:
- White Blood cell count
- Band neutrophils
- Hematocrit
- Platelet Count
- Blood gas measurements (PaO<sub>2</sub> and PaCO<sub>2</sub>)
- pH

## Comparison Summary

<table>
<thead>
<tr>
<th></th>
<th>Physiscore (Our tool)</th>
<th>APGAR (Standard of care)</th>
<th>CRIB</th>
<th>SNAP-II</th>
<th>SNAPPE-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time from birth</td>
<td>3 hours after birth</td>
<td>5 mins</td>
<td>12 hours</td>
<td>12 hours</td>
<td>12 hours</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.91</td>
<td>0.69</td>
<td>0.85</td>
<td>0.82</td>
<td>0.87</td>
</tr>
<tr>
<td>Invasive testing</td>
<td>☒</td>
<td>☒</td>
<td></td>
<td>☒</td>
<td>☒</td>
</tr>
</tbody>
</table>

THE POSITION OF
HISTOLOGY IN THE PROGNOSIS OF
CARCINOMA OF THE BREAST.

By D. H. PATEY, M.S. Lond., F.R.C.S. Eng.,
SURGICAL REGISTRAR, MIDDLESEX HOSPITAL;

AND

R. W. SCARFF, M.R.C.S. Eng.,
ASSISTANT PATHOLOGIST, BLAND-SUTTON INSTITUTE OF PATHOLOGY,
MIDDLESEX HOSPITAL.

[APRIL 21, 1928  801
THE LANCET,]
Cancer Grading System

Results of Investigation.

If the cases are subdivided into groups according to the histological index of malignancy determined as above, there are:---

GROUP 1.—Slight histological malignancy, 16 cases.
Alive and well • • • • • • 11
Died of recurrence • • • • • • 2
Died following operation • • • • • • 1
Died of other causes • • • • • • 2

GROUP 2.—Moderate histological malignancy, 12 cases.
Alive and well • • • • • • 6
Died of recurrence • • • • • • 5
Died of other causes • • • • • • 1

GROUP 3.—Marked histological malignancy, 22 cases.
Alive and well • • • • • • 5
Died of recurrence • • • • • • 17

Thus there does appear to be a progressive deterioration in the results as the histological index of malignancy increases. This supports the conclusions of Greenough, who found:—

Group 1 • • • • 68 per cent. of cases were cured.
Group 2 • • • • 33 " " " "
Group 3 • • • • 0 " " " "

This same basic system is used today!
What information is hiding here?

Break Up Image into Superpixels

Classify superpixels

- Two main types of cells:
  - Epithelial cells – cancer
  - Stromal cells – surrounding tissue
- Train classifier to distinguish superpixels

(Classifier: epithelium vs stroma)

Define broad range of features

- Characterize epithelial and stromal cells
- Spatial relationships
- Contextual features

Acquired H&E stained tissue microarray images from breast cancer patients from 2 institutions:
- Netherlands Cancer Institute (NKI, n=248)
- Vancouver General Hospital (VGH, n=328)

The NKI data set was limited to women less than 53 years with Stage I or Stage II breast cancer.
The VGH data set represents a population-based cohort with a higher proportion of older women and women with more advanced disease.
Building a prognostic model

- Prognostic model trained to the binary target of 5 year survival.
- Algorithm L1-regularized logistic regression
- NKI dataset was used for building the model
  - Performance within NKI assessed by 8-fold cross-validation
- VGH dataset was used for model validation

Prognosis of 5-year survival

(C-Path 5YS Model on the NKI Cohort (n=248))

(A) Prob overall survival

P<0.001

181 153 100 27 4 Low-risk (black)
67 46 29 10 1 High-risk (red)

0 5 10 15 20 25 years

(B) Prob overall survival

P<0.001

239 190 148 95 21 7 Low-risk (black)
89 54 38 21 4 1 High-risk (red)

0 5 10 15 20 25 30 years

Significant stratification even within grade

Grade only weakly correlated with outcome in VGH dataset

11 most predictive features

(probability of overall survival)

$p = 0.006$

Presence of stromal objects without nuclei

Average relative border of stromal spindled nuclear objects to stromal round nuclear objects

Epithelial Features

Stromal features not used at all in current pathological analysis

Stratification within grade

“The field of cancer research has largely been guided by a reductionist focus on cancer cells and the genes within them—a focus that has produced an extraordinary body of knowledge.

Looking forward in time, we believe that important new inroads will come from regarding tumors as complex tissues in which mutant cancer cells have conscripted and subverted normal cell types to serve as active collaborators in their neoplastic agenda.

The interactions between the genetically altered malignant cells and these supporting coconspirators will prove critical to understanding cancer pathogenesis and to the development of novel, effective therapies.”
Data Driven Medicine

- Huge amounts of medical data are now being collected every day
- Much of the previous work is hypothesis-based: testing specific, human-constructed hypotheses
- Unbiased, data-driven analysis can discover novel and important signatures
  - Short-term variability in neonate heart-rate
  - Importance of stromal tissue
- Can be used for prognosis, treatment guidance
- Can provide new insights on disease processes