Detection of Cervical Cancer from Evoked Tissue Fluorescence Images Using 2- and 3-way Methods

Barry M. Wise and Jeremy M. Shaver*
Eigenvector Research, Inc.
Manson, WA

*Now with Just Biotherapeutics
Cervical Cancer

• Pap smears credited with reducing cervical cancer mortality by detecting pre-cancerous cells, but…
• Sensitivity of Pap smears reported as 29-56%
• Abnormal Pap smear -> colposcopic examination, but….
• Colposcopy success depends on interpretation and therefore experience of examiner
• Colposcopic impressions correlate with biopsies as little as 35% of the time
Goal of this work

• Develop a method to classify cervical tissue
  • More sensitive and specific than current methods
  • Doesn’t require high level of experience to use
  • Can be easily administered
The Cervix
Images of the Cervix

- vagina
- cervix
- os
The Data

• Colposcopic Images
  • Interpreted by experts
  • Areas of tissue type identified
• Biopsies
  • Tissue type confirmed using staining and microscopy
  • Areas identified on images
  • “Gold Standard”
• Evoked Tissue Fluorescence Images
  • Excitation Emission Fluorescence Images
Classes of Tissue

- 1-within normal limits
- 2-normal squamous
- 3-normal columnar
- 4-squamous metaplasia
- 5-low SIL (Squamous Intraepithelial Lesion)
- 6-high SIL
Similarity of Tissue Types

- 1-within normal limits
- 2-normal squamous
- 3-normal columnar
- 4-squamous metaplasia
- 5-low SIL
- 6-high SIL
Colposcopic Images and Biopsies

Biopsy locations
# The ETF Images

- Combinations of
  - 3 excitation wavelengths
  - 9 emission wavelengths
  - 22 combinations measured

<table>
<thead>
<tr>
<th>Excitation</th>
<th>337</th>
<th>380</th>
<th>460</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emission (nm)</td>
<td>392</td>
<td>417</td>
<td>452</td>
</tr>
</tbody>
</table>

- Not Measured
Preprocessing Issues

- Image alignment
  - Measurements take about 60s
  - Patient movement an issue
- Patient to patient variability
Image Alignment

• Images at different wavelengths look different
• ...but sub-images should be correlated
  • Should be most correlated when properly aligned
  • Want big PCs to get bigger and small ones to get smaller
• Used Varimax criteria on singular values:
  \[
  \text{Varimax} = \frac{\sum_{i=1}^{n} s_i^4}{\left(\sum_{i=1}^{n} s_i\right)^{4}}
  \]
Variance Captured Before and After Alignment
Model Development

- Align images
- Center to normal squamous tissue on each patient / OR center to mean of all tissues
- Pool all patients - center & scale
- OPTIONALLY: GLS deweighting based on a single class
Calibration

• PLS-DA on EACH CLASS
• take predicted y for each class and
  • threshold
  • convert levels of disease to SIL scale
Application

• Align
• Center each patient image to its own mean (NOTE: high levels of SIL will bias)
• Apply model
• Identify absolute “normal”, repeat (1)-(2) using centering to NORMAL
Validation on Biopsy Regions: Observed probability on each of the 5 PLSDA models

Normal tissues dissimilar

Three disease states very similar!
Diagnostic False-Color Images

Class 2 Model → Normal Squamous Probability
Class 3 Model → Normal Columnar Probability
Class 4 Model → Squamous Metaplasia Probability
Class 5 Model → Low SIL Probability
Class 6 Model → High SIL Probability

Squamous Metaplasia Finder
Diagnostic False-Color Images

Class 2 Model → Normal Squamous Probability

Class 3 Model → Normal Columnar Probability

Class 4 Model → Squamous Metaplasia Probability

Class 5 Model → Low SIL Probability

Class 6 Model → High SIL Probability

High-SIL Finder
Squamous Met. Finder - Classes: 4,2,3

**Red = Squamous Metaplasia**

Hi-SIL Finder - Classes: 6,2,3

**Red = High SIL**

Normal Tissue Finder - Classes: 2,3

**Red/Green = Normal Tissue**

red, green, blue
Key to ROC Plots

Sensitivity / Specificity (at crossing)

n: = results for class “n” model

cv = cross-validation results

val = validation results

modeled class (i.e. 2 vs 3-6)

n: = results for class “n” model

cv = cross-validation results

val = validation results

sensitivity / specificity (at crossing)
Multivariate Curve Resolution
(On pseudo first-order data)
Parallel Factor Analysis
(On second-order data)

![Graphs showing emission and excitation wavelengths.](image)
Summary

• ETF based device very close to clinical usefulness
• Mis-classifications tend to be on progression of disease
• Pre-processing critical
• PLS-DA effective
• Issues
  • Only translational motion considered in alignment
  • Other preprocessing and DA methods to consider