Identifying Constituent Spectra Sources in Multispectral Images To Quantify and Locate Cervical Neoplasia

Kevin C. Baker and Shabbir Bambot

Guided Therapeutics, Inc. 5835 Peachtree Corners East, Suite D Norcross, GA 30092

kbaker@guidedinc.com

INTRODUCTION

- **Current methods** to detect and manage cervical pre-cancers often **miss the disease** and **generate false positives**.
  - This can lead to a delay in correctly diagnosing and over-treating cervical dysplasia.
  - Recent clinical trials with two year follow up have shown that colposcopy (i.e. current triage standard of care) can be inaccurate in determining the need for biopsies and locating appropriate biopsy sites.

- **Guided Therapeutics** has developed a **spectroscopic device** to **reduce** the number of **false positives** while **retaining** the **high true positive** predictive rate.

Figure 1. Spectroscopy measurement device

  - Both reflectance and fluorescence spectroscopy are utilized to detect morphological and biochemical abnormalities associated with cervical pre-cancer and cancer.
  - In a seven-center pivotal study, the device demonstrated its potential as a cost effective and efficient modality for the early detection of moderate and severe cervical intraepithelial neoplasia (CIN) disease in women at risk for cervical disease, while at the same time potentially reducing the number of colposcopies and biopsies from normal and benign cervices.

- **Traditional** supervised and unsupervised **methods** of analyzing fluorescence and reflectance spectra data have **difficulty recovering physiological information.**
  - Supervised un-mixing methods require knowledge about the reflectance/fluorescence spectral patterns of the constituent materials.
  - Reflectance and fluorescence spectral data are non-negative by nature, and spectral sources don't always appear to be statistically independent (e.g. collagen and elastin fluorescence).
  - However, principal component analysis assumes that the underlying sources are uncorrelated and Gaussian, and the resulting sources/concentrations can be negative.
  - Also independent component analysis assumes that the sources are statistically independent, and the sources/concentrations can be negative.

- Therefore, **non-negative matrix factorization** (NNMF) is chosen to analyze the spectral data, since it only **assumes** that the **source** and **mixture data** are **non-negative.**

METHODS

1. **Clinical colposcopist expert and biopsy results are used to create a point-by-point diagnosis.**
   - The diagnosis categories include: normal (e.g. normal squamous, normal columnar), CIN 1 (Grade 1), CIN 2 (Grade 2), CIN 3 (Grade 3), or CIN 2+ (either CIN 2 or CIN3).

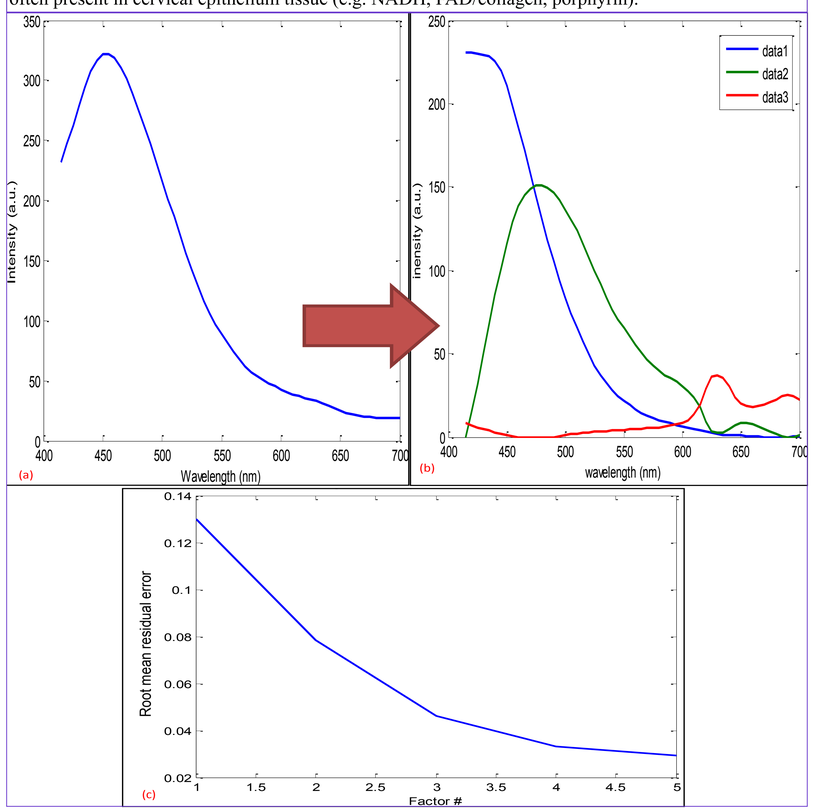
2. **Measurement points and subjects are excluded if they were taken over non-cervical tissue, had excessive blood or mucus, had a specular reflection artifact (CCD saturation), or had an unknown/uncertain disease classification.**

3. **Spectral measurements are calibrated to account for any instrument and subject differences.**
   - Calibration sources are used to calibrate the system's spectra response and perform wavelength calibration.
   - Subject differences are calibrated by dividing the spectra with the mean intensity over all wavelengths.

4. **A NNMF approach is used to blindly decompose reflectance and fluorescence spectra into a set of constituent source spectra curves and concentrations.**
   - NNMF approximates a non-negative matrix with the product of two other non-negative matrices: **X**≈**A***S** where **X** is the measured spectral data matrix, **S** is an unknown spectral source matrix, and **A** is an unknown mixing matrix.
   - Matrices **S** and **A** are chosen to minimize the root-mean-squared residual (RMSR) between **X** and **A***S**.
   - NNMF algorithm is an iterative approach and does not reach a unique solution.

Figure 2. Non-negative matrix factorization approach to construct constituent source spectra.
(a) shows a fluorescence source signal obtained using a single NNMF source model. (b) shows the calculated NNMF source signals obtained using a 3 source model. (c) shows how the RMSR decreases until it becomes stable as the number of factors (sources) included in the model increases.

With a 3 source model, the **source** spectral **signals** appear to **represent physiological fluorophores** often present in cervical epithelium tissue (e.g. NADH, FAD/collagen, porphyrin).



5. **Lasso regression machine learning modeling is used to identify the combination of different source concentrations that best predict the amount of cervical dysplasia.**
   - The Lasso is a shrinkage and selection method for linear regression that minimizes the sum of squared errors with a bound on the sum of the absolute values of the coefficients.
   - The quantitative prediction labels used in the Lasso model are 1 for normal points, 2 for CIN 1, 3 for CIN 2, and 4 for CIN 3.
   - The point dataset is randomly subdivided into 60% training and 40% testing data to test for over fitting.
   - A receiver operating characteristic curve (ROC) analysis is done to determine the dysplasia tissue classification performance.

6. **Two dimensional (2-D) disease maps are created to spatially locate and quantify cervical dysplasia tissue using the Lasso regression results obtained from the NNMF concentrations.**
   - Each subject's 2-D false color disease map is created from the Lasso regression results and the corresponding measurement location.
   - Finally, the 2-D disease maps are compared with cervical colposcopy images and biopsy results.

RESULTS

Figure 3. Comparison between derived and actual source reflectance and fluorescence spectra.
(a) shows the reflection NNMF source signal that best predicts the dysplasia by itself. (b) shows the best fluorescence source signal that predicts cervical dysplasia by itself. (c) shows a hemoglobin (Hb) and oxy-hemoglobin (HbO_2) absorption spectrum that was measured by Scott Prahl at the Oregon Medical Laser Center. (d) shows an actual porphyrin fluorophore spectra signal (i.e. protoporphyrin) at the same excitation wavelength as (b) that was measured by DaCosta at the University of Toronto.

NNMF sources appear to **characterize physiological** cervical tissue **spectra**.
- The behavior of the reflectance spectra is opposite of the absorption spectrum, and you can see the characteristic HbO_2 absorption peaks as valleys in the reflectance NNMF source signal.
- Porphyrin fluorescence spectrum has two peaks near 640nm and 700nm, and these two peaks appear to be also present in the fluorescence NNMF source signal.

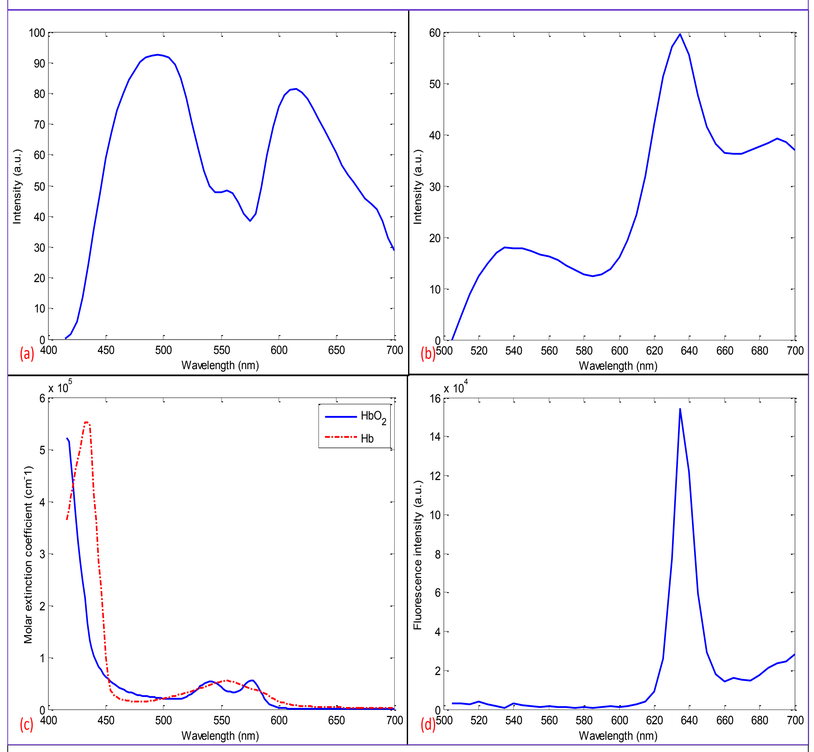
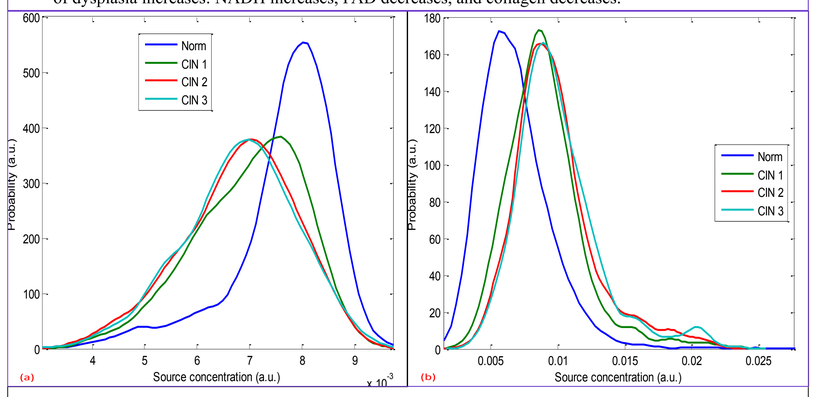


Figure 4. Probability distribution function (PDF) estimates for normal and dysplastic tissue.
- The PDF curves represent the following different cervical tissue: normal (blue), CIN 1 (green), CIN 2 (red), and CIN 3 (cyan). (a) shows the PDF estimates using the reflectance source concentrations that correspond to the Hb source signal shown in Figure 3 (a). (b) shows the PDF curves using concentrations obtained from the fluorescence NNMF signal that appears to be a porphyrin spectrum shown in Figure 3 (b).

NNMF source **concentrations** behave **similar** as **organic compound concentrations** in the cervix.
- Figure 4 (a) results agree with the widely accepted knowledge that as dysplasia increases, the amount of angiogenesis increases and thus the HbO_2 absorption (reflectance) increases (decreases).
- Higher porphyrin concentrations are well known to be present in pre-cancererous tissue.
- Other prevalent cervical tissue fluorophore concentrations (not shown here) behave as expected as the level of dysplasia increases: NADH increases, FAD decreases, and collagen decreases.



ACKNOWLEDGEMENTS

The project described was supported in part by a grant award from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

Figure 5. Lasso probability distribution functions and receiver operating characteristic curves.
(a) shows the PDF curves of the normal, CIN 1, CIN 2, and CIN 3 groups indicating higher values best predict dysplasia tissue. (b) shows classification ROC curves to distinguish CIN 1 and CIN 2+ from normal tissue.

Lasso linear regression **combines** NNMF source **concentration information** to **improve** cervical dysplasia **predication** and **classification performance**.
- There are 18 different features selected after performing the Lasso regression procedure on only the training data.
- Testing dataset had a fitting error that is less than 0.1% different than the training dataset.
- ROC area under the curve (AUC) for CIN 1 and CIN 2+ classification is 0.75 and 0.86
- Detection specificity at 95% sensitivity (spec@95) for CIN 1 and CIN 2+ tissue is 30% and 58%.
- Sensitivity at 70% specificity (sens@70) for CIN 1 and CIN 2+ tissue is 70% and 87%.

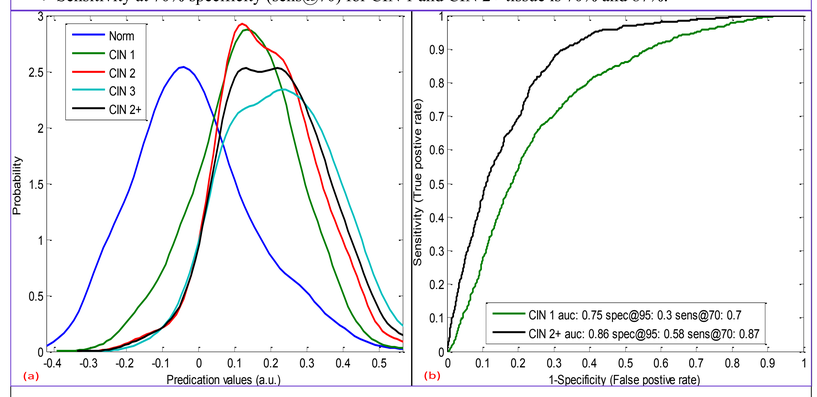
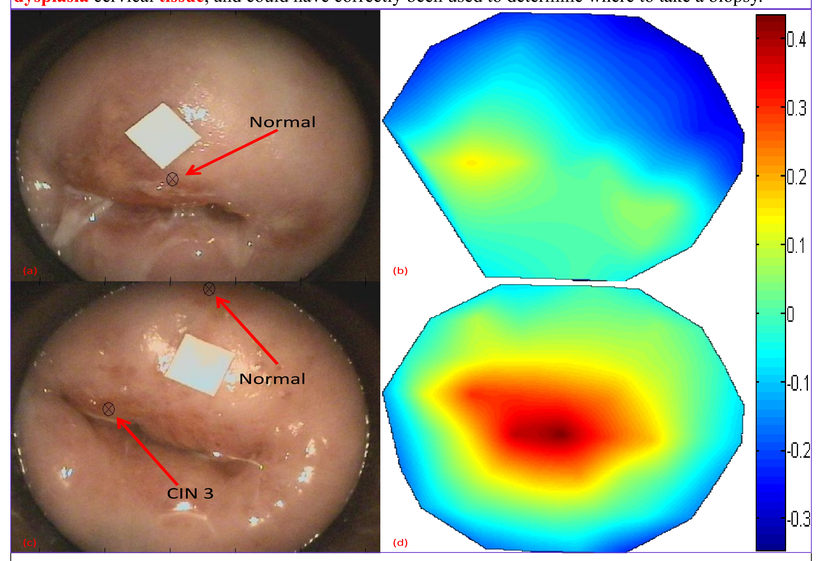


Figure 6. Comparison between colposcopy image and biopsy with 2-D spectroscopy disease maps obtained from the Lasso regression results.
Biopsy spatial locations are shown in the colposcopy images (left images) by an "x" indicator. (a) Colposcopy image with biopsy diagnosed as normal. (b) 2-D spectroscopy disease map showing entire cervix as normal. (c) Colposcopy image with biopsy diagnosed as CIN 3 and normal. (d) 2-D spectroscopy disease map indicating where cervical dysplasia tissue exists.

The corresponding **2-D disease maps** Figure 6 (b) and (d) **correctly identify** the **biopsied normal** and **dysplasia** cervical **tissue**, and could have correctly been used to determine where to take a biopsy.



CONCLUSIONS

- Calculated **sources** from the **NNMF** approach appear to be **represent physiological sources** (e.g. hemoglobin and porphyrin) and their **concentration** changes are in **agreement** with cervical **dysplasia changes**.

- A machine learning algorithm (i.e. **Lasso** linear regression) **improved** the **dysplasia prediction** performance by combining source concentrations from both the reflectance and fluorescence spectra data.

- 2-D false color spectroscopy **disease maps** demonstrate the ability to **quantify** and **spatially locate dysplasia** cervical tissue.

- These results offer the potential to **reduce** the number of **false positive** cases while **maintaining** a high enough **detection rate** necessary for primary screening, and may **assist in identifying biopsy locations**.

REFERENCES

Jeronimo, J., Shiffman, M., "Colposcopy at a crossroads," Am. J. Obstet. Gynecol. 195, 349-353 (2006).

Twiggs, L., B., Chakhtoura, N., A., Werner, C., L., Griffith, W., F., Flowers, L, C., Lashgari, M., Ferris, D., G., Winter, M., L., Sternfeld, D., R., Burnett, A., F., Wilkinson, E., J., and Raab, S., S., "Multimodal Spectroscopy As A Triage Test For Women At Risk For Cervical Neoplasia: Results Of A 1,607 Subject Pivotal Trial," J. Lower Gen. Tract. Dis. 14(3), 264-269 (2010).

Chang, K., S., Mirabal, N., Y., Atkinson, N., E., Cox D., Malpica, A., Follen, M., and Richards-Kortum, R., "Combined reflectance and fluorescence spectroscopy for in vivo detection of cervical pre-cancer," J. Biomed. Eng. 10(2), 024031 (2005).

Hastie, T., Tibshirani, R., and Friedman J., [The Elements of Statistical Learning: Data Mining, Inference, and Predication], Springer, New York NY, (2009).

Berry, M., W., Browne, M., Langville, A., N., Pauca, V., P., and Plemmons, R., J., "Algorithms and application for approximate nonnegative matrix factorization," Computation Statistics and Data Analysis, 52(1), 155-173 (2007).

DaCosta, R., S., Andersson, H., and Wilson, B., C., "Molecular fluorescence excitation-emission matrices relevant to tissue spectroscopy," Photochem Photobiol. 78(4), 384-392 (2003).

http://omlc.ogi.edu/spectra/hemoglobin/

Chang, S., K., Marin, N., Follen, M., and Richards-Kortum, R., "Model based analysis of clinical fluorescence spectroscopy for in vivo detection of cervical intraepithelial dysplasia," J. Biomed. Eng. 11(2), 024008 (2006).

Wu, Y., Xi, P., and Qu, J., Y., "Depth-resolved fluorescence spectroscopy of normal and dysplastic cervical tissue," Optics Express 13(2), 382-388 (2005).