

In-vivo Detection of Cervical Cancer Lesions using Hyperspectral Colposcopy

Carlos Vega^{1,*}, Norberto Medina², Raquel Leon¹, Himar Fabelo^{1,3,4}, Alicia Martín², Gustavo M. Callico¹

¹Research Institute for Applied Microelectronics (IUMA), University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain.

²Complejo Hospitalario Universitario Insular Materno Infantil (CHUIMI), Servicio Canario de Salud (SCS), Las Palmas de Gran Canaria, Spain.

³Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Las Palmas de Gran Canaria, Spain.

⁴Research Unit, Hospital Universitario de Gran Canaria Dr. Negrin, Las Palmas de Gran Canaria, Spain

*cvega@iuma.ulpgc.es

Supplementary Results

Multiclass classification for CIN lesion grading

This experiment aims to demonstrate, as a proof of concept, the potential of the proposed method to spatially diagnose at pixel level, the cervical region in different classes. The multiclass classification method is trained for grading the different CIN lesions established in the Bethesda system¹, where it is possible to classify each pixel into healthy, LSIL, HSIL or IC. **Fig. S1** presents the validation results for the three HS data modalities: *reflection* (**Fig. S1a**), *absorbance* (**Fig. S1b**), and *second derivative* (**Fig. S1c**). These results show more noticeable differences than the HSIL+IC classifier, in which the second-derivative mode did not yield accurate results for the HSIL class across most classifiers. Results in the reflectance and absorbance modes were similar, but the reflectance mode had a slightly lower mean standard deviation across the 4 classes (0.27) compared to the *absorbance* mode (0.25), which is not statistically significant. The best performing classifier is the SVM-RBF-C in the reflectance mode, which achieved an F1-score of 0.62 ± 0.06 , and the IC was the best performing class with a score of 0.67. This model was then retrained using all training and validation data before being evaluated on the independent test set.

The quantitative results presented in **Fig. S1d-f** for F1-score, precision, and recall, respectively, show that the healthy class are noticeably higher than the results for the other classes, with an F1-score of 0.50 ± 0.28 compared to the 0.05 ± 0.01 , 0.15 ± 0.28 , and 0.21 ± 0.17 from the LSIL, HSIL, and IC, respectively. The comparison between mean spectral signatures and standard deviations across training and test groups (Supplementary Fig. S2) reveals spectral variations among patients that could limit the classifier's generalisation. In addition, the overlap between some classes is evident: LSIL in the training group closely matches healthy cervical tissue in the test group (Fig. 4g), and HSIL shows substantial similarity to IC, leading to label swaps in some patients.

Representative examples from different diagnostic grades are presented in **Fig. S1h** for a qualitative analysis. From these patients, *P4* was diagnosed as CIN3 and manually labelled with a combination of HSIL and LSIL areas, however, the classifier identifies the whole region as IC. This contrasts with its heatmap for the HSIL class (which includes the CIN3 grade), where the lesion area indicated in the ground truth shows a high response for this class. This suggests that the HSIL and IC classes produced similar outputs, with the HSIL class showing a slightly lower score. Additionally, *P13* is an IC case, but most annotated pixels are classified as HSIL, with a high probability in the heatmap. This discrepancy strongly penalises the classification results, even though both pathologies require immediate treatment from a clinical perspective.

Supplementary Figures

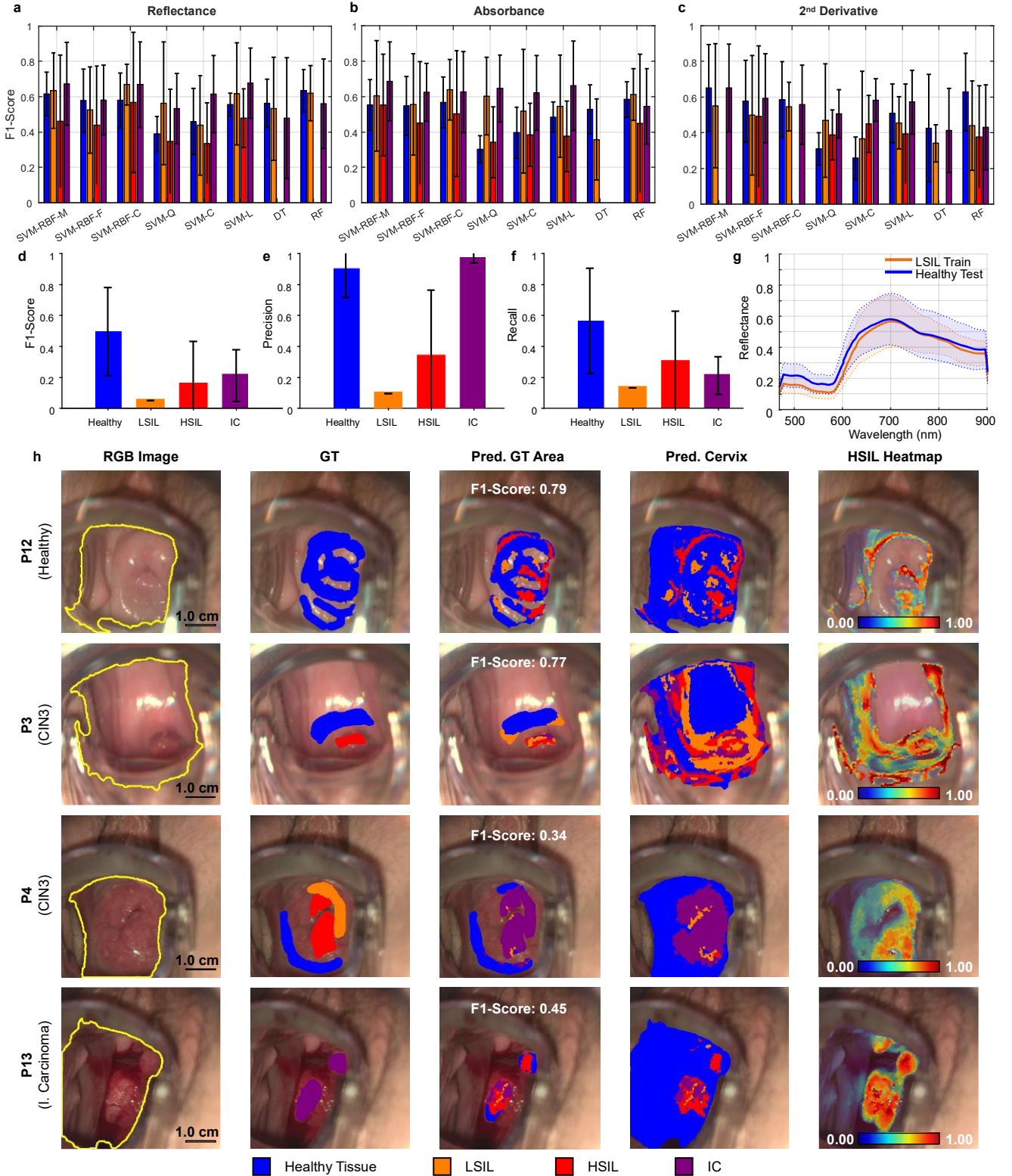


Fig. S1| Multiclass classification results and quantitative performance metrics. a-c, Validation results reported as mean performance across folds for the different classifiers and the three HS data modalities: a, reflectance; b, absorbance; and c, second derivative. Bars show means; error bars denote standard deviation across folds. d-f, Quantitative evaluation on the independent test set: d, F1-score; e, precision; f, recall. g, Representation of mean reflectance and standard deviation for the LSIL class in the training patients (orange) and the healthy class in the test patients (blue). h, Representative examples of pixel-wise classification for different diagnostic patients. For each patient (left to right): synthetic RGB; prediction for the GT pixels (healthy in blue, LSIL in orange, HSIL in red, IC in purple); prediction for the whole cervical region; and the heatmap for HSIL class (scale 0–1). Yellow contours delineate the cervix; scale bars, 1.0 cm.

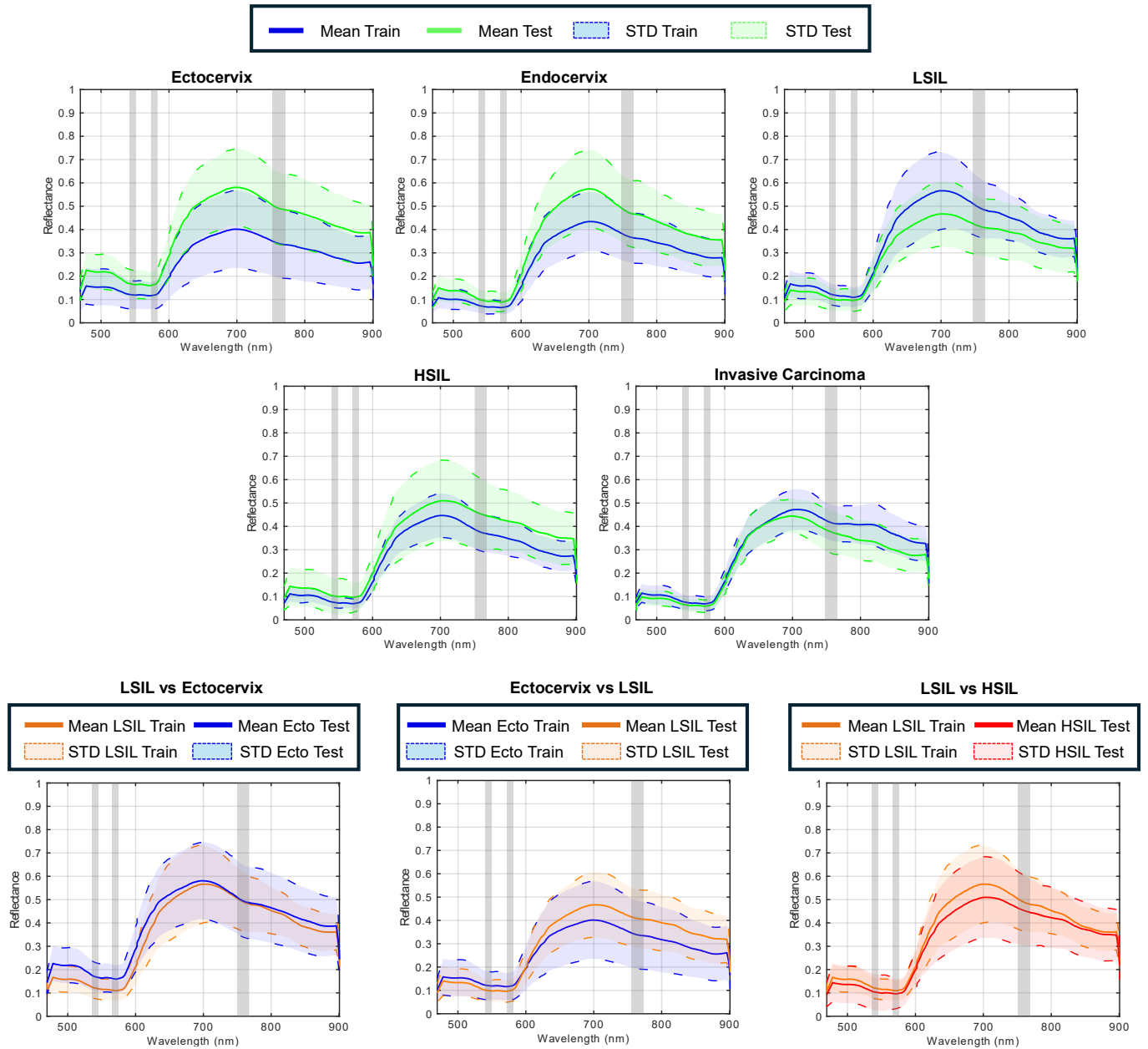


Fig. S2| Mean spectral signatures for each diagnostic class in the training versus test cohorts. For each graph shaded areas indicate standard deviation across pixels. Differences between cohorts and overlaps between class means (for example, ecto vs. LSIL; LSIL vs. HSIL) illustrate potential sources of generalisation error. The darker grey areas represent the wavelengths related to the most relevant absorption peaks of haemoglobin in 546, 576 and 750 nm.

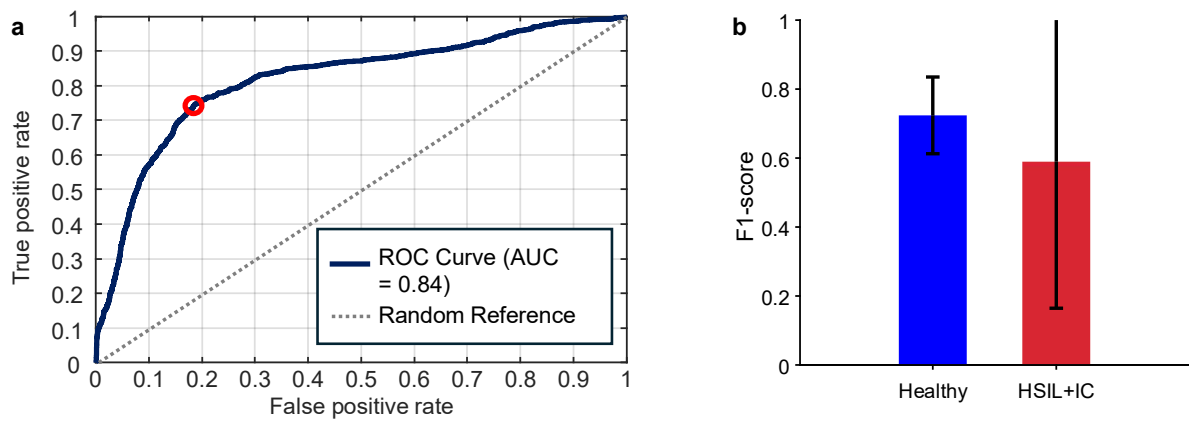


Fig. S3 | ROC curve and class-wise performance for HSIL+IC classification in the test group. a, Receiver operating characteristic (ROC) curve of the binary classifier for discriminating HSIL+IC from healthy cervical tissue in the independent test set. The ROC curve was obtained by varying the decision threshold applied to the classifier's output score for the HSIL+IC class and calculating the corresponding true and false positive rates at each threshold. The area under the curve (AUC) from this ROC curve was 0.84. The red marker indicates the selected operating point used for classification, and the dotted diagonal line represents random performance. b, Class-wise performance of the binary classifier in the independent test set for healthy tissue and HSIL+IC. Bars represent the mean score for each class, and error bars indicate the standard deviation across the test cohort.

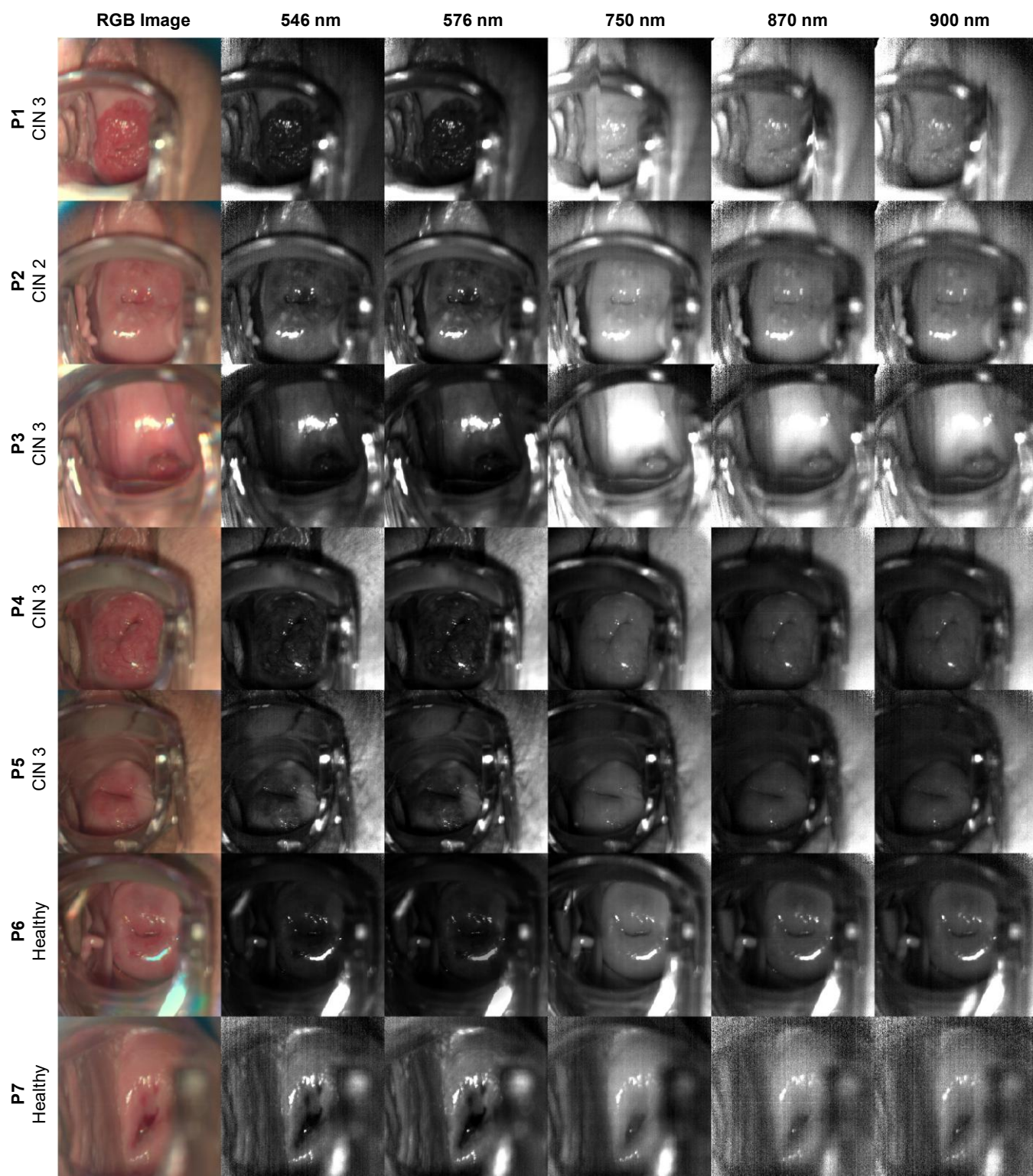


Fig. S4| Hyperspectral captures for test patients P1–P7. For each patient (rows), the panel shows a synthetic RGB composite reconstructed from the HSI cube, followed by single-band images at 546, 576, 750, 870 and 900 nm

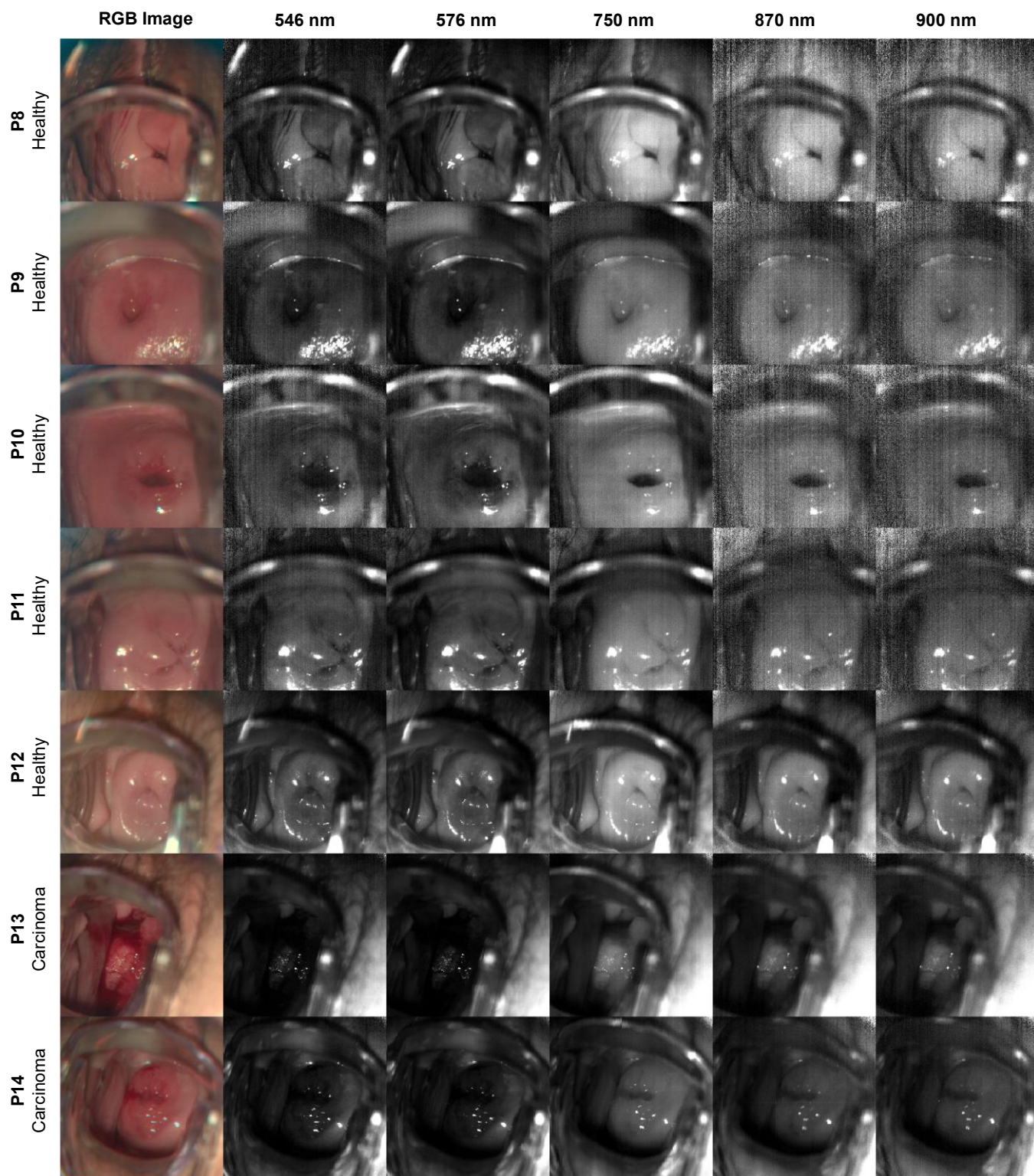


Fig. S5 | Hyperspectral captures for test patients P7–P14. For each patient (rows), the panel shows a synthetic RGB composite reconstructed from the HSI cube, followed by single-band images at 546, 576, 750, 870 and 900 nm