Development of algorithms for automated detection of cervical pre-cancers with a low-cost, point-of-care, Pocket Colposcope

Mercy Nyamewaa Asiedu[1,3,4]*, Anish Simhal[2], Usamah Chaudhary[5], Jenna L. Mueller[1,3,4], Christopher T. Lam[1,3,4], John W. Schmitt[6], Gino Venegas[7], Guillermo Sapiro[1,2,7], Nimmi Ramanujam[1,3,4]

[1]Department of Biomedical Engineering, Duke University, Durham, North Carolina, United States of America

[2]Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, United States of America

[3]Center for Global Women's Health Technologies, Duke University, Durham, North Carolina, United States of America

[4]Duke Global Health Institute, Duke University, Durham, North Carolina, United States of America

[5]Department of Computer Science and Department of Mathematics, Duke University, Durham, North Carolina, United States of America

[6]Department of Obstetrics and Gynecology, Duke University, Durham, North Carolina, United States of America

[7]Liga Contra el Cáncer, Pueblo Libre, Lima, Peru

*Corresponding author

Email: mercy.asiedu@duke.edu

Abstract

The World Health Organization recommends visual inspection with acetic acid(VIA) and/or Lugol's iodine(VILI) for cervical cancer screening in low-resource settings. Human interpretation of diagnostic indicators from visual inspection is qualitative, subjective and has high inter-observer discordance, which could lead both to adverse outcomes for the patient and unnecessary follow-ups. In this work, we propose methods for (1) automatic feature extraction and classification for VIA and VILI cervigrams and (2) combining features of VIA/VILI cervigrams for improved performance. Cervix images (cervigrams) were acquired with a low-cost, miniature, digital colposcope. We describe algorithms to pre-process pathology-labeled cervigrams and extract simple but powerful color and textural-based features. The features are used to train a support vector machine(SVM) to classify cervigrams based on pathology for VIA, VILI, and combination of the two contrasts. The proposed framework achieved a sensitivity, specificity, and accuracy of 81.3%, 78.6%, and 80.0%, respectively when used to distinguish cervical intraepithelial neoplasia (CIN+) relative to normal and benign tissues. This is superior to the average values achieved by expert physicians on the same data set for discriminating normal/benign from CIN+ (sensitivity=77%, specificity=51%, accuracy=63%). The results suggest that utilizing simple color- and textural-based features from VIA and VILI images may provide unbiased automation of cervigrams, enabling automated, expert-level diagnosis of cervical pre-cancer at the point-of-care.

Introduction

In the United States and other high-income countries, cervical cancer incidence and mortality have decreased by 70% over the last 70 years [1, 2]; however, women living in low and middle-income countries (LMICs) still experience a disproportionately high burden of cervical cancer [3, 4]. In fact, about 85% of the global burden of cervical cancer occurs in LMICs, with 87% of cervical cancer related deaths occurring in these regions [5]. In 2012, over 500,000 women were diagnosed with cervical cancer, and approximately 50% of them died from this disease [5]. If status quo is maintained, the numbers are expected to rise to over 750,000 new cases and 400,000 deaths per year by 2035 [5].

In high-income countries, screening for cervical cancer is performed using a multi-tiered paradigm which begins with the Papanicolaou (Pap) smear with human papillomavirus (HPV) co-testing, followed by colposcopy guided biopsy and treatment if needed [6]. Colposcopy uses a low-power microscope to visualize cervix changes that are highlighted with the application of exogenous contrast agents, such as acetic acid and in some cases Lugol's iodine. Acetic acid (3% or 5%) application to the cervix causes a reversible coagulation in nuclear proteins and cytokeratin, which primarily affect lesion areas due to high nuclear proteins content [7]. This causes whitening and mosaic textured features in abnormal regions, whereas normal cervix regions remain a light pink color. Normal areas of the cervix are glycogen rich and take up Lugol's iodine, which turns normal cells dark brown, while abnormal areas are glycogen deficient and do not uptake the glycophilic Lugol's iodine solution and hence, appear pale/mustard yellow [7]. Suspicious areas are biopsied for pathology confirmation of cervical abnormalities [6]. Women with pre-cancer are treated via excision of a portion of the cervix using Loop Electrosurgical Excision Procedure (LEEP). Women with cancer are referred to a combination of local and/or systemic therapy depending on the stage of invasive disease. The screening model is not practical to implement in LMICs due to the resources that are required to procure and implement well-established solutions [8, 9].

In LMICs, HPV testing has been recommended as an alternative to the Pap smear. However, HPV testing is not widely available in LMICs [8], and if available, requires a secondary test that provides specificity in order to prevent large numbers of women from being unnecessarily referred for confirmatory testing or treatment [8]. Visual inspection with acetic acid (VIA) using the naked eye is a low-tech version of colposcopy

that serves to either add specificity to HPV testing (when available) or serves as the primary screening tool. However, VIA is not scalable for a number of reasons. Specifically, the lack of image capture results in poor quality control due to lack of documentation, and lack of magnification diminishes sensitivity to microscopic disease [10-16].

Our group has developed a low-cost, portable, point-of-care digital colposcope, called the Pocket Colposcope, to replace VIA with high-quality colposcopy, which is the established standard-of-care for triage in high-resource settings [17-19]. The Pocket Colposcope has the form factor of a tampon with a 5 megapixel (MP) color Complementary metal–oxide–semiconductor (CMOS) camera and light emitting diodes (LEDs) that provide white and green light illumination. Image capture and device power is provided through USB connectivity to a cell phone, tablet, or computer [17, 19]. An international study conducted by our group found high concordance between the Pocket Colposcope and standard-of-care clinical colposcopes [20]. The Pocket Colposcope can overcome the limitations of VIA in settings where traditional colposcopes are too cumbersome and/or expensive to use. This will not only provide enhanced visualization and magnification compared to traditional VIA but will also provide much needed documentation and feedback to ensure quality control in healthcare provider performance. However, the limited availability of expert colposcopists in low resource settings represents yet another bottleneck [8, 9]. Therefore, our goal is to develop image processing and machine learning-based approaches for images obtained with the Pocket Colposcope to potentially recapitulate the expertise of an experienced colposcopist in settings where there is a shortage of experts available to screen large populations.

Several groups working on algorithms for automated computer-aided colposcopy have shown proof-of-concept image analysis and classification to automatically interpret acetowhitening for cervical cancer detection [21-43]. These methods have been predominantly applied to VIA cervigrams only, which do not fully capture all sources of contrast described in the modified Reid's Colposcopic Index (mRCI) [7]. Additionally, most of these methods require pre-selection of suspicious regions (semi-automated), and have not been demonstrated to be scalable, particularly since these methods still rely on subjective user interpretation. Visual inspection with Lugol's iodine (VILI), which provides another important source of contrast for the visualization of cervical abnormalities, has not been evaluated previously. Studies with physicians providing diagnosis

based on both contrasts have shown that VILI has the potential to bolster the performance of algorithms based on VIA alone [18, 44]. However, other studies have found that while the parallel combination of VIA and VILI bolsters sensitivity, it reduces specificity [45].

Our goal is to develop a series of feature extraction and simple machine algorithms that leverage both VIA and VILI images obtained with Pocket Colposcope for the automated diagnosis of cervical pre-cancers to recapitulate expert colposcopist performance. By developing a novel strategy incorporating VILI algorithms, we hypothesize we can achieve improved sensitivity and specificity performance over VIA alone. Unlike previous approaches, this method does not require a health provider to pre-select an area of concern, but rather evaluates the entire cervix to automatically identify regions of interest. The algorithms work by pre-processing images to reduce specular reflection, automatically segmenting a region of interest from the cervix for analysis, extracting color- and texture-based features, and using a support vector machine for binary classification of VIA and VILI images. Receiver operating characteristics (ROC) curves are generated from the classifier to determine the area under the curve (AUC), which indicates how well a model predicts classes [46]. VIA and VILI algorithms were then combined using two methods: (1) in parallel, where VIA and VILI image pairs were processed with their respective algorithms and scores were used to generate a classifier model and (2) serially, where VIA images were run through the VIA algorithm and then VILI images of the VIA positive results were run through the VILI algorithm. With the proposed framework, VIA only, VILI only, parallel and serial combined algorithms' sensitivities were 70.3%, 64.1%, 79.7%, and 81.3%, respectively; specificities were 72.9%, 84.3%, 80.0%, and 78.6%, respectively; and overall AUCs were 0.79, 0.79, 0.86, and 0.86, respectively compared to biopsy-confirmed pathology. As a point of comparison, expert physicians' interpretation achieved an average sensitivity of 77% and specificity of 51% for the same dataset.

Methods

Overview

First, VIA and VILI images from the Pocket Colposcope cropped to remove clinically irrelevant features such as the speculum, vaginal walls and specular reflection automatically attenuated. For both VIA and VILI images, a Gabor filter was applied to automatically segment regions within the cervix for further analysis based on color and texture differences. Haralick's features (contrast, correlation, energy and homogeneity) [47] were

calculated from these segments. The segments were also transformed to different color spaces (grayscale, RGB, YCbCr, HSV and CIElab). From each color space central tendencies, mean, median, mode, variance and Otsu threshold level [48] were calculated for each color channel of each color space. Additionally, for VILI images the percentage of pixels corresponding to the yellow staining from non-uptake of Lugol's iodine by the lesion region was used to determine a pseudo lesion size. An optimal subset of these features was selected using a wrapper forward sequential feature sub-selection for the individual VIA and VILI algorithms. These features were used to train a support vector machine (SVM) classifier using pathology as ground truth labels, and 5-fold cross validation was used to determine performance and generate ROC curves for binary classification of images as VIA/VILI positive or negative. For the parallel-combined algorithm, scores for classifying the VIA and VILI images with their respective algorithms were input as predictors to generate a classifier for combined classification. For the serial-combined algorithm, the VIA algorithm with selected features was first applied to the VIA image data set at a threshold set to enable high sensitivity which resulted in high positivity rate. VIA negatives were classified as negatives by the algorithm while VIA positives were further analyzed with the VILI algorithm. Corresponding VILI images of the VIA positives were analyzed with the VILI algorithm (which has higher specificity) for binary classification as negative or positive. Results for VIA only, VILI only and the combined methods were compared to each other and to the performance of expert physicians on the same data set. Figure 1 provides a summary of the steps of the individual VIA and VILI algorithms described, which is further described below. It is important to note that the algorithm uses a combination of domain-knowledge inspired features with machine learning, since unlike other more standard applications of machine learning, including that in the medical field (e.g.., viz.ai), the amount of training data is limited. However as shown here, not fully needed when such smart features are incorporated.

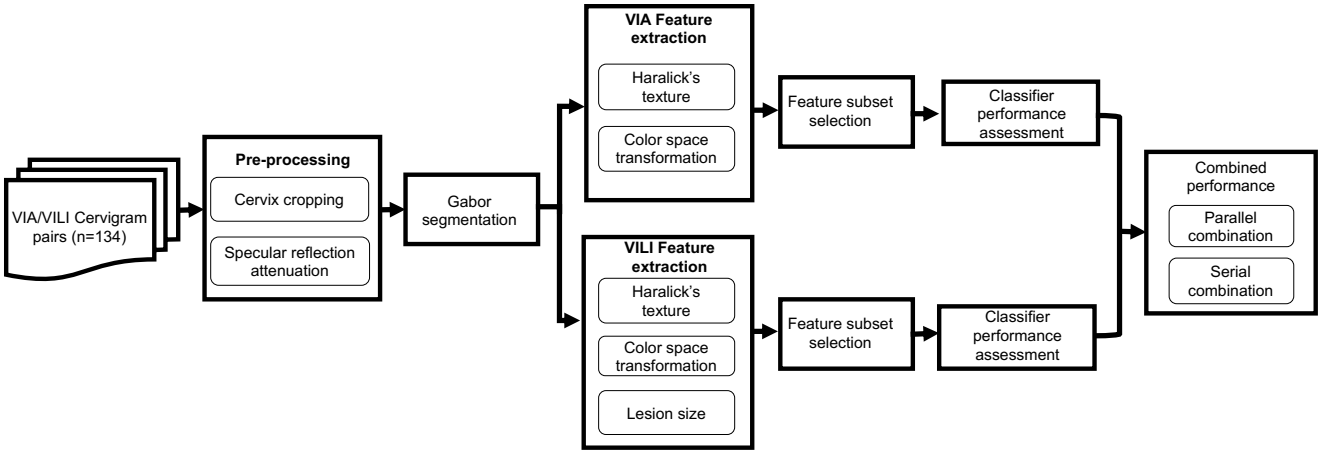

Figure 1. Flow chart of individual feature extraction and classification process. Process shows VIA/VILI cervigram processing to remove clinically irrelevant features, Gabor segmentation, feature extraction, feature subset selection, classifier training and validation and combination of VIA and VILI for classification.

Image collection

Images and labels were retrospectively obtained from a database of Pocket Colposcope images acquired in a previous clinical study [20] in which blinded expert physicians provided a diagnosis for each patient from reviewing randomized digital cervigrams based on both VIA and VILI images. To summarize, two hundred patients undergoing colposcopy examination at La Liga Contra el Cancer Clinic in Lima, Perú were recruited for a clinical study. This study was approved by Duke University institutional review board (Pro00052865) and performed with approved protocol, informed consent process, and data storage system at La Liga Contra el Cancer, Lima, Peru. For each patient, acetic acid was applied to the cervix and images were captured with the standard-of-care colposcope followed by the Pocket Colposcope. Lugol's iodine was then applied to the cervix and VILI images captured with the standard-of-care colposcope followed by the Pocket Colposcope. Images, patient demographics, and pathology results were collected and stored in a HIPAA compliant secured database REDcap [49]. No biopsies were taken for normal colposcopies as per standard-of-care procedures. Hence the ground truth for VILI negative images was primarily based on expert physician interpretation. Of biopsies taken based on positive colposcopy, 6 came back normal, 19 came back as benign conditions (cervicitis and condilomas) and the remainder were pre-cancers (CIN1, CIN2+) or cancers (invasive carcinomas). For algorithm binary classification labels, condilomas and cervicitis were considered pathologically normal. 25 out of 51 (49%) colposcopy-negatives were pathology-confirmed negative and 26 out of 51 (51%) were negative based on interpretation by the lead colposcopist at the clinic. VIA/VILI image pairs

which were deemed interpretable by an expert physician were randomized and sent out to three expert colposcopists for cervical pre-cancer diagnosis. Physician interpretation was based on both VIA and VILI features and were classified as either VILI/VIA positive or negative.

Cervix pre-processing

Cervix cropping

Cervigrams typically contain clinically superfluous features such as the vaginal sidewalls and the speculum. These artifacts can affect feature extraction and diagnostic accuracy of an automated algorithm and therefore were removed to enable accurate analysis and classification of features pertaining to the cervix. Due to image positioning diversity which for each cervigram, the cervix region of interest (ROI) was cropped using a minimum bounding box around the cervix region identified by an expert colposcopist. With standardized images in which the cervix took up about 90% of the image, no cropping was necessary.

Specular reflection attenuation

Specular reflections appear as bright white spots on the cervigram where there is saturation of light exceeding the camera's linear detection range for a set exposure. Specular reflections from Pocket Colposcope images were primarily caused by moisture, the uneven surface of the cervix, excess light illumination, and for the VILI images light reflecting off the dark VILI stains. Specular reflection affects subsequent processing methods, which are primarily color-based, hence the need to attenuate their effect. We employed the well-established specular reflection attenuation method described by Das et al. (2010) [50]. In summary, the raw RGB image was separated into the R, G, and B channels. Specular reflection was identified by finding pixels above a threshold (220 for an 8-bit image) from each channel and logically AND-ing them. In the binary image, the outlines of reflection were dilated, and the borders of dilated outlines were identified on the original image. A Laplacian infill method using values from these borders was used to smoothly interpolate inward from the border pixels [50], providing an output image with reduced specular reflection.

Feature extraction

Gabor segmentation

A Gabor algorithm combined with k-means clustering was applied to the grayscale of the raw RGB image to select a region within the cervix for further analysis. Gabor filters look for specified frequency content in an

image in specified regions and directions, similar to the human visual system [51]. Our hypothesis was that the Gabor filter would have a high frequency response within the acetowhite/texturized regions in VIA positive and yellow lesion regions in VILI positives. K-means clustering, an unsupervised learning method which splits unlabeled data into k number of clusters [52], was then used to segment the Gabor filtered image into two clusters, and the cluster with the higher Gabor mean selected for further analysis. For VIA/VILI negative regions, the assumption was that a region within the cervix would be selected for further analysis, which would subsequently indicate negativity. Utilizing this approach overcomes limitation from previous studies where regions needed to be manually selected for further textural analysis. A multi-scale Gabor filter was created with different orientations (0, 45, 90, 135 degrees) and different wavelengths in increasing powers of two starting from $4\sqrt{2}$ up to the hypothenus length of the input image. The orientations and wavelengths used in this study are well established for unsupervised texture segmentation [51]. The filter was convolved with each image to obtain the overall Gabor response. K-means clustering (k=2) was used to segment the cervix into two regions based on the Gabor response. The region with the highest mean Gabor response was selected for further analysis. Gabor segmented regions were used to calculate color, and Haralick's texture features were used for both VIA and VILI images. Pseudo lesion size was also calculated for VILI images.

Haralick's texture features

Since acetic acid application causes textured mosaicism, which highlights different vasculature patterns corresponding to different diagnosis, we calculated Haralick's texture features [47] to analyze the Gabor-segmented VIA regions. This method was also applied to VILI images with the assumption that since VILI is applied after VIA, there may be texturized regions within the VILI segments. In the greyscale image, we calculated the grey-level co-occurrence matrix (GLCM), a second order statistical method which characterizes an image by calculating how often a pair of pixels with specific values and spatial relationship occur in an image. The GLCM was computed for four different pixel offsets (1, 5, 10 and 15) each in four different directions (0, 45, 90, 135 degrees), for a total of 16 matrices. Small to medium offset values were used since we expected mosaicism features to show pixel differences within this range. The different directions were used since features of interest may be horizontal, vertical or diagonal. From these GLCMs, four Haralick's features

were calculated: contrast, correlation, energy and homogeneity were calculated. This yielded 4 texture-based features for each VIA and VILI image.

Color space transformation calculations

According to the mRCI, both VIA and VILI cervigram classification are primarily based on staining. For VIA, acetowhitened regions correspond to lesions, while light pink regions correspond to normal epithelium. For VILI, yellow regions correspond to lesions, while darker brown staining corresponds to normal epithelium. Based on this biological effect explained in detail in the introduction, color-based features were extracted. Each image was transformed to five different color spaces and these separated into individual color channels, specifically 'Red(R), Green(G), Blue(B) (RGB)', 'grayscale', 'Hue(H), Saturation(S), Value(V) (HSV)', 'Luminance(Y), blue chroma(Cb) and Red chroma(Cr) (YCbCr)', and 'Lightness(L), red/green(a), and yellow/blue(b) (CIELAB)' [53]. RGB color space is widely used for digital image acquisition; however, one of its primary disadvantages is that it is device dependent and sensitive to illumination differences, which limits the use for color-based image analysis [54]. CIELAB color space extends the RGB color space from 90% to 100% of all perceivable colors and enables more quantitative applications since it correlates to perceptual descriptors of color [55]. Additionally, CIELAB and the YCbCr color spaces separate image luminous intensity from color information, reducing factors resulting from how an image was taken, such as variations in LED brightness [53]. HSV rearranges the geometry of RGB color space to make it more intuitive and perceptually relevant, separating color based of its hue, saturation and lightness [53]. Gray scale of an RGB image has luminance information of the image [53].

For each color channel, we determined summary central tendencies of mean, median, mode, and variance. Otsu threshold level (two-class classification) was also determined on color-space transformed images [56] to take advantage of multiple clusters present from the brown normal regions and yellow abnormal regions. The Otsu threshold is a widely used, nonparametric, unsupervised method, which automatically selects a threshold to separate an image into two clusters, maximizing inter-class variance and minimizing intra-class variance [48]. From each of the thirteen channels of the different color space transforms, the mean, median, mode, variance, and Otsu's threshold were extracted as potential classification features. Overall 65 color-based features were extracted for each VIA and VILI image.

Lesion size determination

A pseudo lesion size detection was performed for VILI images after Gabor segmentation. Due to the high variation in acetowhitening lesion and normal epithelium shades, it was not possible to determine lesion size for VIA images based on acetowhitening alone. Since VILI images provide a higher, more consistent contrast between dark brown normal epithelium and yellow stained abnormal epithelium, color-based “pseudo lesion size” was determined using VILI images. VILI images were transformed into CIElab color space explained previously and the b channel was extracted due to its direct linearity of yellow shades and high contrast in color transforms. A threshold was determined via analysis of pixel intensities of ROIs from a subset of images corresponding to lesions and non-lesion regions. The number of pixels above this threshold was divided by the total number of pixels in the bounded cervix region to approximate a percent “lesion size”. We call it pseudo lesion size because Lugol’s iodine solution may also result in yellowing due to non-uptake in benign cervical conditions, such as condilomas, cervicitis. Parts of the cervix where Lugol’s Iodine is not properly applied (such as cervix boundaries) also show non-uptake yellow features. This resulted in overall a total of 70 features for VILI and 69 features for VIA images.

Feature selection, classifier training and testing

Classifier design

We selected a support vector machine classifier, a widely used and successful supervised learning method, which outputs an optimal hyperplane for classification [53, 57, 58]. A labeled training data set is input into the classifier which then builds a model to enable classification of a new data set. For non-linearly separable data sets, the SVM can transform data points to higher dimensions to linearize them for optimal separation using higher order kernels [59]. From initial scatter plots demonstrating non-linearity of the data set, we selected a Radial Basis Function (RBF or Gaussian) kernel with automatically selected kernel size using cross validation. RBF kernel was selected because it is most widely used due to flexibility and ability to project data into high (infinite) dimensionality space to find a linear separation [59]. The features were used as predictors while the pathology diagnosis (normal and benign vs. CIN+) was used as input labels for binary classification. While other machine learning tools could potentially be used, SVM was selected due to the limited data available for training and its excellent performance as reported below.

Feature selection and validation

In machine learning, classifying data using a proportionally high number of features compared to sample size can result in classifier failure due to redundancy and overfitting or overtraining to noise/artifacts. High feature number also increases storage and computational requirements. Due to a limited sample size (n=134 VIA/VILI pairs), a small subset of important features for satisfactory classification performance was selected using forward sequential feature selection (FSFS) with minimum classification error (MCE) [60]. The FSFS (Figure 2) is a fast, wrapper feature selection method, in which feature candidates are sequentially added to an empty set until there is no further improvement in prediction. For each feature subset, the function performs a 5-fold cross validation with the classifier. The feature number and combination, which provided the minimum error for classification is then selected.

Training and testing

The optimal feature subset was used as predictors into the SVM to generate a classification model which was validated using 5-fold cross validation with pathology as ground truth (note that pathology is not available even to clinicians doing visual observation, so this ground truth is a very high standard). For cross validation for VIA and VILI features, the data set was divided into 5. Training was performed on 4/5 of the data and testing performed on the remaining 1/5 of the data. This was repeated leaving a different data set each time and the results averaged. ROCs were generated to determine the AUC, and sensitivity and specificity of each the algorithm was compared to physician performance for classifying images as negative or positive.

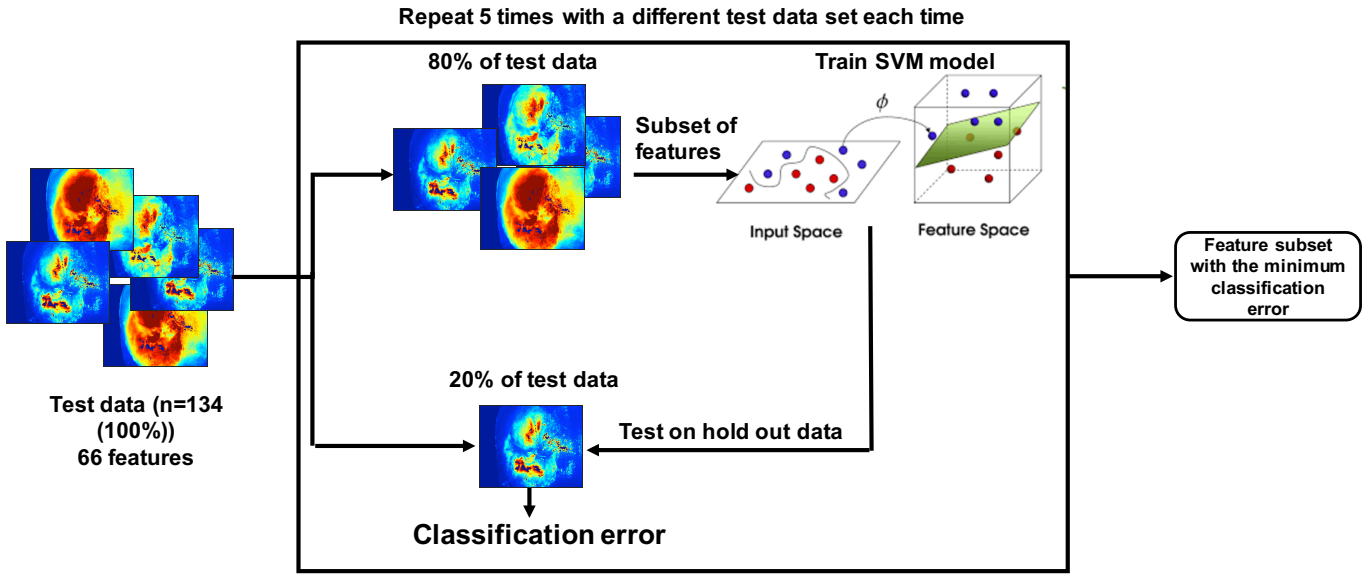

Figure 2. Summary of feature sub selection using wrapper method forward sequential forward sequential feature selection.

Combining VIA and VILI features

Parallel combination

When the SVM classifier classifies VIA and VILI images as negative and positive it also assigns a score to each image which describes the probability or likelihood that the label comes from one of the classes. These probabilities for VIA and VILI images were used as predictor inputs into the parallel-combined SVM classification model with binary pathology labels. A classification model was then generated from these to provide binary classification based on both VIA and VILI inputs.

Serial combination

For serially combining VIA and VILI results, we analyzed the data from the individual VIA and VILI algorithms to determine where they complemented each other. Based on this analysis, we first ran the VIA algorithm, using the same feature subset on all images with a low classification cutoff (-0.25 instead of 0) for binary classification. This enabled classification of all high grades but also resulted in a high false-positivity rate. Negatives based on this classification were considered true negatives. The VILI algorithm was then run on the VIA positive images to increase specificity. An SVM classifier model was generated for the VIA and then VILI stage and cross-validated with pathology as ground truth.

Results

Image breakdown

Of the 200 Pocket Colposcope patients, 66 were excluded due to one or more of the following factors: screen fail (2.5%), missing pathology for ground truth (2%), images not saved (missing) (6%), blurriness (10%), and device set at the wrong working distance (7%). Device related issues have been addressed in a newer version of the Pocket Colposcope design and improved control software. For the purpose of this study, data from a total of 134 patients were used for image analysis. The pathology distribution of patient whose cervigrams were utilized for the study are outlined in Table 1.

Table 1: Image pathology distribution (n=134)

VIA/VILI classification	Pathology classification	# of patients (%)
Negative	Normal	51 (38%)
	Cervicitis	17 (13%)
	Condiloma	2 (0.01%)
Positive	CIN1	41(31%)

	CIN2	10(0.07%)
	CIN3	5(0.04%)
	Invasive Cancer	8(0.06)

Image processing and feature extraction

Specular reflection attenuation

Pre-processing to attenuate specular reflection was performed prior to feature extraction and classification to prevent inaccuracies due to glare. Figure 3 shows results from representative images with extreme cases of specular reflection before and after specular reflection attenuation. Due to the speckled nature of specular reflection it is difficult to accurately hand select all glare regions for quantitative comparison and analysis of reduction. However, upon visual inspection all images had significantly reduced specular reflection without affecting contrast in the images. This finding is similar to previous work in this area [50].

Gabor segmentation

A Gabor filter was applied to pre-processed VIA/VILI images and the response normalized. Lesion regions for both VIA and VILI Images had higher Gabor responses than the surrounding epithelium. Images were segmented into two main regions by k-means clustering (k=2) and the region with highest mean Gabor response selected. Since VIA and VILI negative images had no or little acetowhitening/yellowing, regions selected were likely to be a subsection of the normal epithelium of the cervix, the os or the transformation zone. From direct observation, the Gabor segments were in approximately the same regions for both VIA and VILI for 130/134 of the images. The Gabor segmented enabled the automatic selection of a region of interest for further analysis with the color and Haralick's texture features.

Haralick's texture features

VIA positive images had significantly high contrast (p=0.03), not significantly different correlation, significantly low energy (p=0.013) and significantly lower homogenity (p=0.02) compared to VIA negative images. This is expected since "contrast" corresponds to variations between pixels (0 for a constant image), correlation looks at a mutual relationship between pixels (range: -1 to 1, with Nan for a constant image), energy crresponds to uniformity, with higher values for similar pizels (0-1 with 1 for a constant image) and homogenity reflects closeness in distribution to the diagonal of GLCM elements (range: 0-1, with 1 for diagonal/constant image), which calculate how often a pair of pixels with specific values and spatial relationship occur in an image. VILI

positive images had significantly higher contrast (p=0.003), insignificantly different correlation, significantly lower energy (p=0.024) and significantly lower homogeneity (0.015) compared to VILI negative images.

Color transforms features

Each image was transformed into 5 main color spaces and the central tendencies and Otsu threshold calculated. Figure 3 shows representative color transforms for the VIA (YCbCr-Cb) and VILI (CIElab-b) images. As shown, lesion regions for VILI tended to have lower values than non-lesion regions in the b channel. In the Cb channel for VIA, lesion regions tended to have higher values than non-lesion regions. Different color channels showed distinct trends as well as redundancies. The optimal combination of features was later addressed using the sequential feature selection method. Representative images for all color channels are shown in the appendix.

Pseudo lesion size detection

Pseudo lesion size detection was performed on pre-processed VILI images since they enabled higher color-based contrast between lesion and non-lesion regions than VIA. We found that a threshold of 40 in the b channel of CIElab color space enabled adequate cutoff between the yellow and brown regions of the VILI images, the percentage of pixels above a threshold of 40 (which corresponds to all shades of yellow) was divided by the total cervical area to compute lesion size. The mean lesion size was found to be significantly higher (p=0.0047) in VILI negatives than in VILI positives as expected.

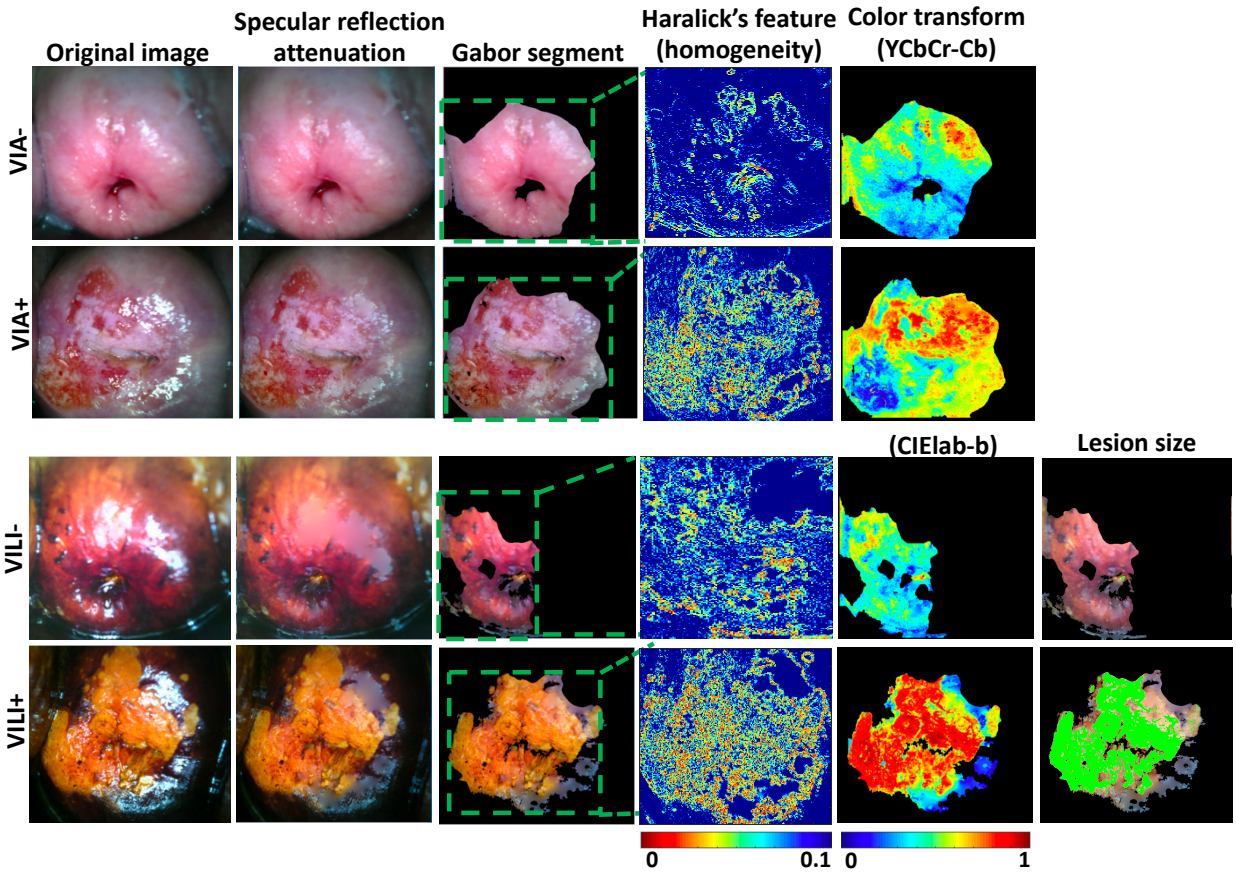

Figure 3. Endpoints for feature extraction methods used for VIA and VILI classification

Feature selection

Feature selection was performed to select a subset of features to prevent overfitting. Rather than use a simple filter method which selects features with the highest p-values, we used a wrapper forward sequential feature selection (FSFS) method which takes into consideration feature redundancies and interaction. The list of selected features and their p-values are in Table 2. Figure 4 compares results of classification scores for VIA and VILI, achieved using features selected from the simple filter selection method and the forward sequential feature selection (FSFS). Results from VIA and VILI images show that features selected with the FSFS method shows expected trends in pre-cancer grades and significantly higher differences and AUC between normal/benign and CIN+ compared to intuitively selected features from the simple filter selection method.

Table 2. Features and p-values from simple filter method and sequential feature sub-selection method.

VIA				VILI			
Features w/top p-values		**Sequential selected**		**Top P values**		**Sequential selected**	
Feature	**P-value**	**Feature**	**P-value**	**Feature**	**P-value**	**Feature**	**P-value**
Correlation	0.0009	A mode	0.07108	CB median	0.0001	B mode	0.0005
Blue variance	0.0016	B mode	0.2077	CB mean	0.0001	Red variance	0.5681
L variance	0.0065	CR mode	0.6073	CB mode	0.0001	CB mean	0.0001
Y variance	0.007	Energy	0.0009	CB level	0.000195	CB level	0.0002
Grey variance	0.007	Correlation	0.90849	H mean	0.00024	Red mode	0.17415
Green variance	0.0147	CR level	0.5717	B mean	0.00042	V mode	0.3329

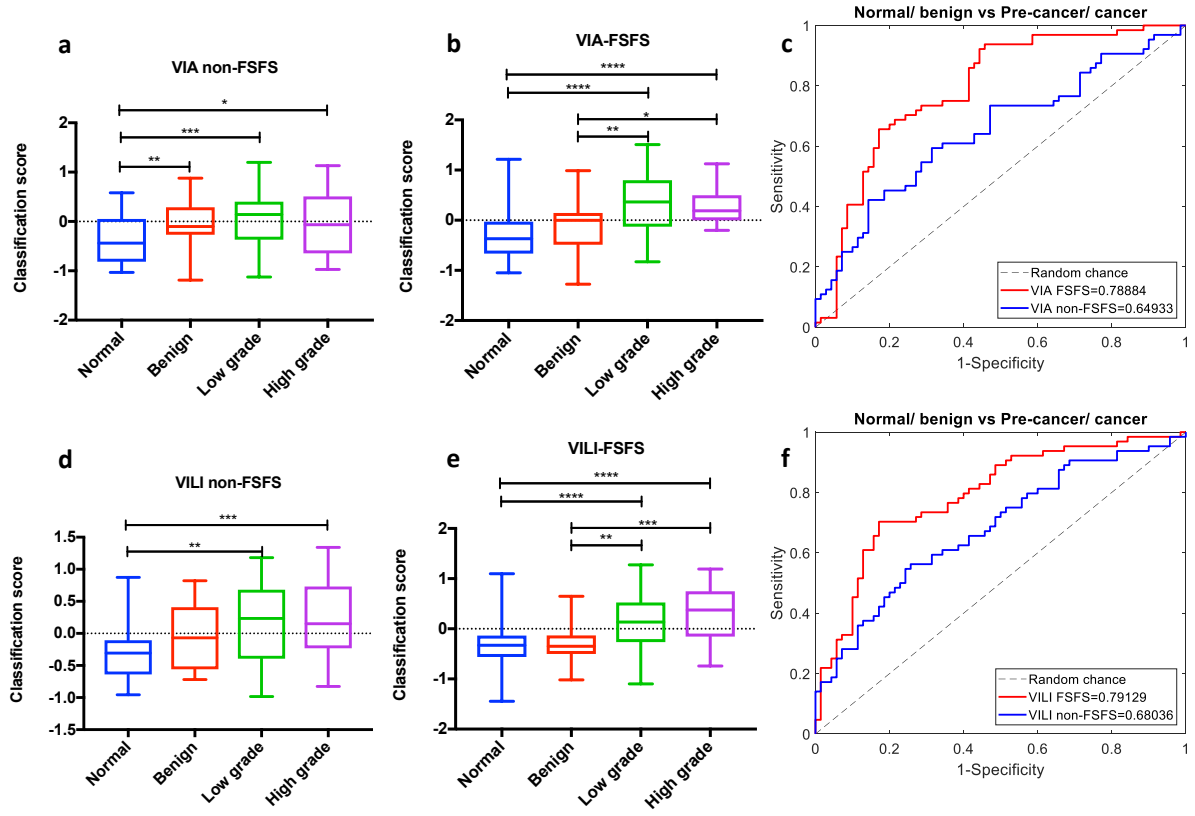

Figure 4. Comparing classifications scores from using top features by p-value (simple filter method) and by FSFS method. a) Box plots of classification scores for VIA algorithm using the features with the highest p-values, b) Box plots of classification scores for VIA algorithm using features selected with sequential feature sub-selection method, c) ROC curves comparing VIA results from features with simple filter method to FSFS method, d) Box plots of classification scores for VILI algorithm using the features with the highest p-values, e) Box plots of classification scores for VILI algorithm using features selected with sequential feature sub-selection method, f) ROC curves comparing VILI results from features with simple filter method to FSFS method. *=p<0.05, **=p<0.001, ***=p<0.0001, ****=p<0.00001.

SVM classification

Figure 5a and b show scatterplots of classification scores for the parallel and serial combination methods. Classification scores are outputs from the SVM classifier indicating the probability/likelihood of an image belonging to either the negative or positive class. Scatterplots show similar expected trends for both methods. ROCs were generated to compare how the different algorithms (VIA, VILI parallel combined and serially combined) performed compared to the pathology ground truth for normal/benign vs. CIN+, normal/benign vs. low grade/CIN1 and normal/benign vs. high grade/CIN2+. Performance of three expert physicians (>15 years' of colposcopy experience per person) for the same data set are also indicated on the curve. The physician interpretations were collected retrospectively and were based on features from both VIA and VILI image interpretation. Hence physicians had the same information as the combined VIA and VILI algorithms.

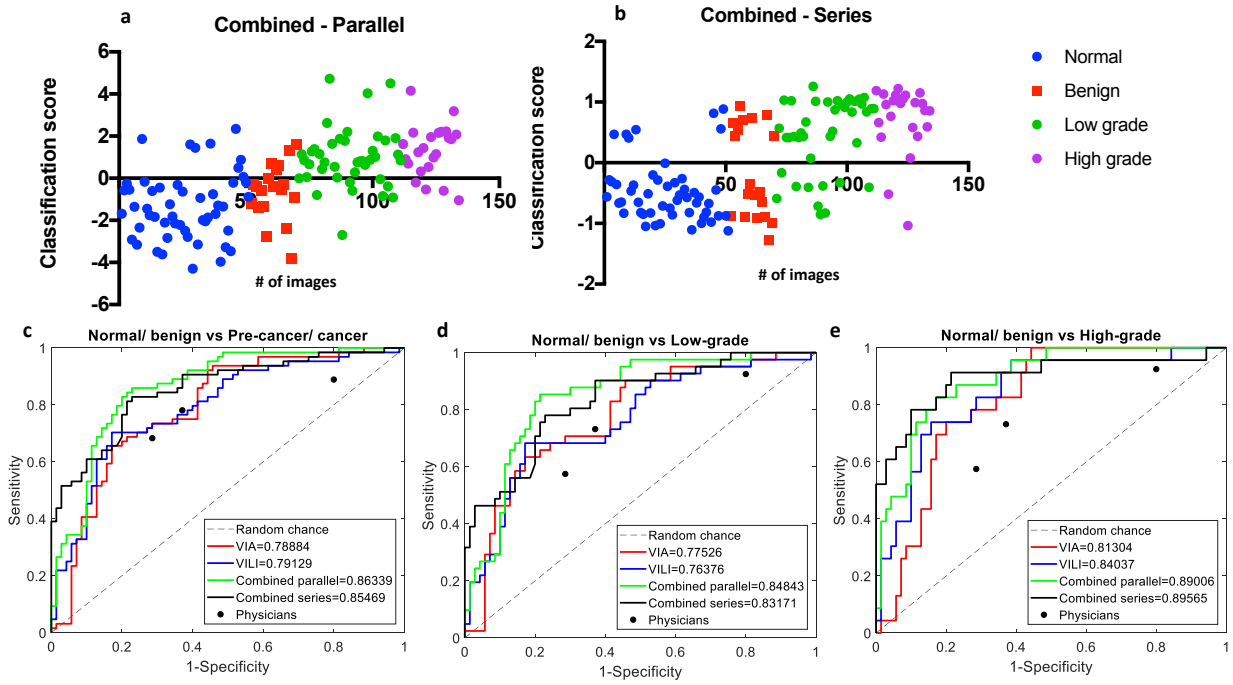

Figure 5. Performance of different classification algorithms. a) Scatterplot for classification scores from parallel combined method, b) Scatterplot for classification scores from serial combined method, c) ROC curves for all algorithms for normal/benign vs CIN+ (d) ROC curves for all algorithms for normal/benign vs low grade/CIN1, e) ROC curves for all algorithms for normal/benign vs CIN2+. Performance of each expert physician reader (n=3) is also indicated for each graph. # of images refers to the number of images/ image number.

All physician performance points fall within the ROC curve for combined VIA and VILI algorithms, indicating superior algorithm performance, which is critical considering experts like the ones used in this study are not widely available and certainly lacking in low resource areas. Both combined algorithms outperformed the algorithms using the single contrasts but achieved similar AUCs. Additionally, AUCs for VIA and VILI only AUCs were also similar. The best performing algorithm achieves an AUC of 0.86 for discriminating normal/benign from all CIN+, 0.85 for discriminating normal/benign from low-grades/CIN1 and 0.89 for discriminating normal/benign from high-grades/CIN2+.

Figure 6 shows algorithm and each individual physician's sensitivities, specificities and accuracy as compared to pathology (Figure 6a and b). Where sensitivity also known as the true positive rate measures the proportion of correctly identified positive, specificity also known as true negative rate measures the proportion of correctly identified negative cases and accuracy measures the overall proportion of correctly identified cases. Overall accuracy for all algorithms are higher than physician accuracies. Though physician's sensitives are on average on par with algorithms, algorithms achieve higher specificities. Combined algorithms achieved highest agreement with physicians, particularly with physicians who had higher performance with pathology. This is predictable since physicians also used both VIA and VILI images for prediction. Figure 6c shows the percentage of correctly classified images by pre-cancer grade for one of the combined (serial) algorithms and the best physician. The algorithm outperforms the best physician across all cervix types but most notably for benign lesions which tend to be overcalled by physicians.

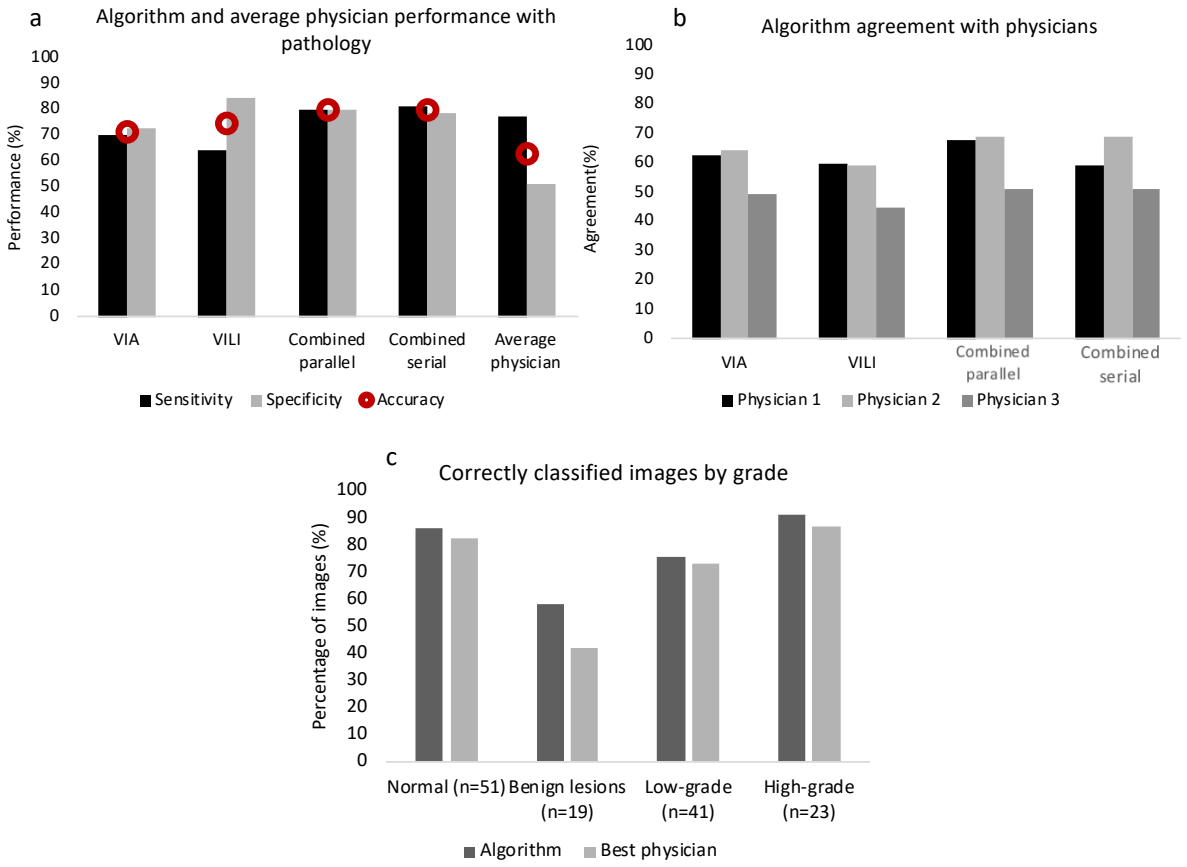

Figure 6. Bar charts for classification performance between algorithm, physicians and pathology. (a) Sensitivity, specificity and accuracy for the 4 algorithms as well as average physician performance compared with pathology. (b) Sensitivity, specificity and accuracy for individual physicians compared with pathology. (c) Comparing number of correctly classified images by pathology grade for combined algorithms and the best physician.

Figure 7 shows representative VIA/VILI cervigrams of true positive, true negatives, false positives and false negative images classified by the serially combined algorithm. Figure 7a-d show VIA images with their corresponding VILI cervigrams in Figure 7e-h. True positive images showed acetowhitening and mosaicism in VIA image (Figure 7a) and yellow staining in VILI image (Figure 7e). True negatives showed little to no visible whitening or mosaicism in VIA positive image (Figure 7b) and showed brown staining from iodine uptake in VILI image (Figure 7f). False positive Images, classified as negative by pathology but positive by the algorithm were due to high texture in the VIA image (Figure 7c) and the non-uptake of Lugol's iodine in the VILI image (Figure 7g). The false negative showed no distinct acetowhitening in the VIA image (Figure 7d) and very little area of Lugol's non-uptake in the VILI image (Figure 7h). This may be a result of the limitation where biopsies are taken from the endocervical canal which is not typically visible in the image. To conclude, these images as

well as all the errors of the automatic algorithms “make sense” and are a result from non-trivial images (recall that we compared with pathology, which is data not available to the algorithm or to the clinician).

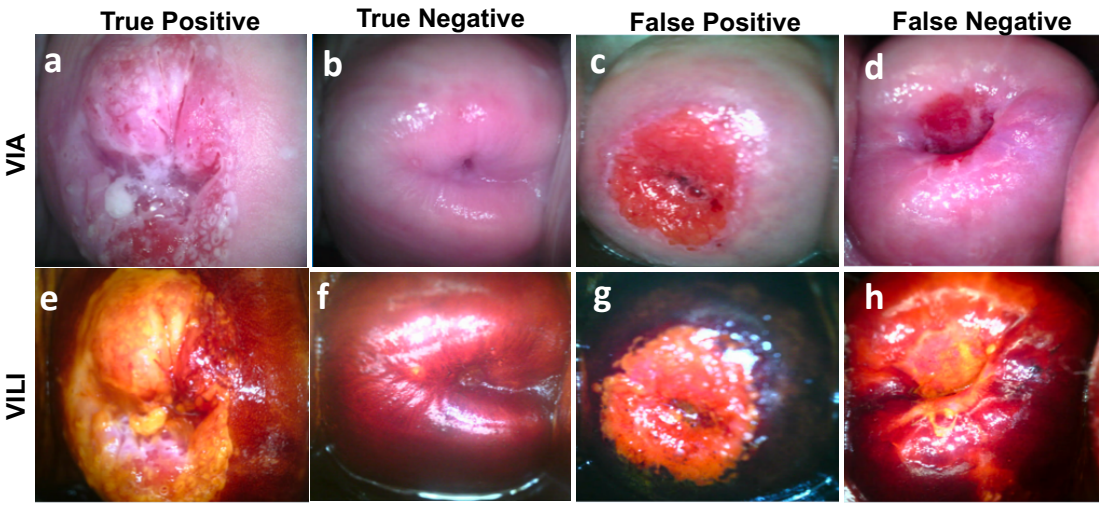

Figure 7. Representative VILI (A-J) and corresponding VIA (E-H) images diagnosed by algorithm. (A,E) True positive, (B,F) True negative, (C,G) False positive and, (D,H) False negative. Physician decisions were made off both VILI and VIA features whereas the algorithm only used VILI features.

Discussion

We introduced in this study algorithms to process and classify VIA and VILI cervigrams as negative or positive for cervical pre-cancer and to combine features from both contrasts to improve overall specificity while maintaining sensitivity. Algorithms for the individual contrasts were developed to extract color and texture-based features. A subset of the domain/expertise-motivated features was automatically selected and used to train a support vector machine to develop models for binary classification of cervigrams. The individual algorithms were combined both in parallel (both at the same time) and serially (VIA followed by VILI) to improve performance. Combined methods performed on par with each other, with the parallel method achieving a sensitivity, specificity, accuracy and AUC of 79.7%, 80.0%, 80.0%, and 0.86, respectively, and the serially combined algorithm achieving a sensitivity, specificity, accuracy, and AUC of 81.3%, 78.6%, 80.0%, and 0.86, respectively. This was on higher than average expert physician, who achieved a sensitivity, specificity and accuracy of 77%, 51%, and 63%, respectively. The algorithm more accurately classified negatives, benign conditions, low grades and high-grade conditions than the best performing physician.

This algorithm has been developed with images captured using the Pocket Colposcope, a low-cost, portable, digital colposcope. The labeled images used for this study were obtained retrospectively and physicians used both VIA and VILI data, hence we were not able to directly compare the individual algorithms for VIA and VILI. No previous algorithms for combining VIA and VILI classification have been developed to which we can compare our results. There was only one study we found that had proposed a method to combine VIA and VILI images for diagnoses [61]. However, though preliminary representative images from color-based segmentation were shown, no quantitative performance of the VILI algorithm was provided to which we could compare our results. Additionally, there is no previous algorithm that uses VILI alone for cervical pre-cancer diagnosis. Several groups have proposed methods towards automatic VIA diagnosis which, include automated cervix segmentation, specular reflection removal, acetowhitening detection using color-based and texture features. Several groups are working to create algorithms for automated computer-aided (CAD) colposcopy and have already shown proof of concept and feasibility that automated image analysis can be used to interpret colposcopy images for cervical cancer detection [21-43]. Their findings can be broken down into data management of uterine cervix images, pre-processing for image enhancement, feature extraction, and lesion detection. However, most of these are only semi-automated, requiring manual ROI selection and focusing on lesion detection within abnormal VIA images [23]. Manually selected regions are then classified as a type of cervix tissue using various classifiers [23]. Even though some groups have started working to enable full automation, the current methods are device- and illumination-dependent with illumination inhomogeneity and image diversity being cited as a major challenge in analysis [41, 62].

We are able to compare physician performance in our study to that from previous studies. A colposcopy study by Qureshi et al with 328 women found a sensitivity of 86.84% and a specificity of 48.93% [63]. A study in Tanzania by Ngoma et al of 10,367 women screened with VILI found a range of 79.8% - 99.3% for sensitivity and 97.0% - 97.6% for specificity [64]. Sankaranarayanan et al performed a study with 4,444 women in Kerala, India, and found a range of 80.6% - 92.0% for sensitivity and 83.6% – 85.8% for specificity [65]. Our physician performance ranged from 68.3 - 88.9% for sensitivity, 20.0 - 71.4% for specificity, and 52.6 - 70.1% for accuracy. It is interesting to note while the average sensitivity of physicians in this study was comparable to that reported in previous studies, the average specificity was lower. We speculate that this may be because

physicians who participated in our study were viewing still images of VIA/VILI without any information on patient history or demographics compared to previous studies in which images were interpreted in real time. Additionally, the prevalence of disease is lower at the population level, which was the study population used in previous studies, compare to the diagnostic population (secondary screening), which was the study population investigated here. In the diagnostic population, there is a higher number of borderline benign conditions, which can be easily misinterpreted due to intrinsic properties of the contrast agents.

Future work includes testing the algorithm in a real time clinical setting and comparing its performance to pathology and expert physician interpretation. To improve algorithm functionality, we will employ registration of VIA and VILI images for classification. We were unable to do so in this study because the camera is taken out between VIA and VILI imaging for application of the contrast. Thus, VIA and VILI images were not guaranteed to be in the same position, which prevents superimposing the images. Future additions to the Pocket Colposcope that will enable standardized image capture could make this possible and would be included in future versions of the algorithm. Standardized image capture will also enable automated cervix cropping or remove the need for cervix cropping since only a minimum bounding box around the cervix region is needed. Additionally, including vascular information could improve specificity. Neo-vascularization is a well-established phenomenon associated with the development of high-grade pre-cancers in the cervix [66, 67]. We have developed algorithms that can segment and quantify blood vessels and demonstrated its utility in a pre-clinical model. Specifically, we designed a dark field microscope and a Gabor filter based vessel segmentation algorithm to quantify early neovascularization changes in the dysplasia-carcinoma sequence [68]. Tortuosity, vessel diameter, and vessel length showed significant contrast between pre-cancers and normal tissues stratified by small vessels (<16 μm diameter) and large vessels (≥16 μm diameter). We have found in preliminary analysis that the Pocket colposcope recapitulates vascular imaging with the dark field microscope. This data suggests that incorporating high-resolution vascular imaging has the potential to improve upon the specificity of acetowhitening and Lugol's iodine contrast.

Overall, the algorithm for automated cervical cancer diagnosis performed significantly better than average diagnosis of expert physicians. However, a limitation of the study was that most of the features were assumed rather than learned. Application of techniques to learn features from images, which we may possibly

be missing, could also improve performance. However, the proposed algorithm requires very little training data, computational space and run time compared to other machine learning frameworks. Even though this algorithm is developed using images captured with the Pocket Colposcope, the method can potentially be applied to cervigrams from other colposcopes.

Conclusion

We have developed a method for image processing of VIA and VILI cervigrams and selected features to train a support vector machine algorithm for classifying images as negative or positive. The algorithms can also be used to provide a diagnosis with combined VIA and VILI features. The algorithm includes automatic specular reflection removal, transformation of the image into different color spaces and channels for automatic feature extraction, and a fully automated classification decision rule, resulting in real time, computer-aided diagnosis of cervical pre-cancer. This method has potential to enable colposcopy diagnosis in the absence of an expert physician or to complement his/her judgment. Combining VIA and VILI classification outperforms physicians, thus allowing accurate referral colposcopy to be more accessible at the primary care/community level setting.

References

1. Group, F.I.S., *Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions.* N Engl J Med, 2007. **356**(19): p. 1915-27.
2. Paavonen, J., et al., *Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women.* Lancet, 2009. **374**(9686): p. 301-14.
3. Scarinci, I.C., et al., *Cervical cancer prevention: new tools and old barriers.* Cancer, 2010. **116**(11): p. 2531-42.
4. Freeman, H.P et al. Excess cervical cancer mortality: A marker for low access to health care in poor communities. National Institutes of Health; 2007. Report No.: 05-5282.
5. (IARC), I.A.f.R.o.C. *GLOBOCAN Cervix Uteri ASR (W) per 100,000, all ages*. 2012 [cited 2016 27 June]; Available from: http://globocan.iarc.fr/old/bar_sex_site.asp?selection=4162&title=Cervix+uteri&statistic=2&populations=6&window=1&grid=1&color1=5&color1e=&color2=4&color2e=&submit=%C2%A0Execute.
6. Massad, L.S., et al., *2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors.* J Low Genit Tract Dis, 2013. **17**(5 Suppl 1): p. S1-S27.
7. Group, I.S. *The modified Reid colposcopic index (RCI)**. 2017 [cited 2017 May 17]; Available from: http://screening.iarc.fr/colpoappendix5.php.
8. Jeronimo, J., et al., *Secondary Prevention of Cervical Cancer: ASCO Resource-Stratified Clinical Practice Guideline.* J Glob Oncol, 2017. **3**(5): p. 635-657.
9. WHO, in *Comprehensive Cervical Cancer Control: A Guide to Essential Practice*, nd, Editor. 2014: Geneva.

10. Sherigar, B., et al., *Cervical Cancer Screening by Visual Inspection with Acetic Acid - Interobserver Variability between Nurse and Physician.* Asian Pacific Journal of Cancer Prevention, 2010. **11**(2): p. 323-326.
11. Raifu, A.O., et al., *Determinants of Cervical Cancer Screening Accuracy for Visual Inspection with Acetic Acid (VIA) and Lugol's Iodine (VILI) Performed by Nurse and Physician.* PLoS One, 2017. **12**(1): p. e0170631.
12. Vedantham, H., et al., *Determinants of VIA (Visual Inspection of the Cervix After Acetic Acid Application) positivity in cervical cancer screening of women in a peri-urban area in Andhra Pradesh, India.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(5): p. 1373-80.
13. Gravitt, P.E., et al., *Effectiveness of VIA, Pap, and HPV DNA testing in a cervical cancer screening program in a peri-urban community in Andhra Pradesh, India.* PLoS One, 2010. **5**(10): p. e13711.
14. Labani, S., et al., *CareHPV cervical cancer screening demonstration in a rural population of north India.* Eur J Obstet Gynecol Reprod Biol, 2014. **176**: p. 75-9.
15. Longatto-Filho, A., et al., *Performance characteristics of Pap test, VIA, VILI, HR-HPV testing, cervicography, and colposcopy in diagnosis of significant cervical pathology.* Virchows Arch, 2012. **460**(6): p. 577-85.
16. Jeronimo, J., et al., *A multicountry evaluation of careHPV testing, visual inspection with acetic acid, and papanicolaou testing for the detection of cervical cancer.* Int J Gynecol Cancer, 2014. **24**(3): p. 576-85.
17. Asiedu, M.N., et al., *Design and preliminary analysis of a vaginal inserter for speculum-free cervical cancer screening.* PLoS One, 2017. **12**(5): p. e0177782.
18. Jenna Mueller, C.L., Denali Dahl , Mercy Asiedu , Marlee Krieger , Yenny Bellido-Fuentes , Mark Kellish , Jenna Peters , Alaattin Erkanli , Ernesto Ortiz , Lisa Muasher , Peyton Taylor , John Schmitt , Gino Venegas , Nimmi Ramanujam, *Portable Pocket colposcopy performs comparably to standard-of-care clinical colposcopy with acetic acid and Lugol's iodine in Lima, Perú.* International Journal of Obstetrics and Gynaecology, 2018 (Under review).
19. Lam, C.T., et al., *Design of a Novel Low Cost Point of Care Tampon (POCkeT) Colposcope for Use in Resource Limited Settings.* PLoS One, 2015. **10**(9): p. e0135869.
20. Mueller, J.L., et al., *International Image Concordance Study to Compare a Point-of-Care Tampon Colposcope With a Standard-of-Care Colposcope.* J Low Genit Tract Dis, 2017. **21**(2): p. 112-119.
21. Garcia-Arteaga, J.D., J. Kybic, and W.J. Li, *Automatic colposcopy video tissue classification using higher order entropy-based image registration.* Computers in Biology and Medicine, 2011. **41**(10): p. 960-970.
22. Gordon, S., et al., *Evaluation of uterine cervix segmentations using ground truth from multiple experts.* Comput Med Imaging Graph, 2009. **33**(3): p. 205-16.
23. Greenspan, H., et al., *Automatic Detection of Anatomical Landmarks in Uterine Cervix Images.* Ieee Transactions on Medical Imaging, 2009. **28**(3): p. 454-468.
24. Ji, Q., J. Engel, and E. Craine, *Classifying cervix tissue patterns with texture analysis.* Pattern Recognition, 2000. **33**(9): p. 1561-1573.
25. Jondet, M., R. Agoli-Agbo, and L. Dehennin, *Automatic measurement of epithelium differentiation and classification of cervical intraneoplasia by computerized image analysis.* Diagn Pathol, 2010. **5**: p. 7.
26. Lange, H., *Automatic glare removal in reflectance imagery of the uterine cervix.* Medical Imaging 2005: Image Processing, Pt 1-3, 2005. **5747**: p. 2183-2192.
27. Lange, H., *Automatic detection of multi-level acetowhite regions in RGB color images of the uterine cervix.* Medical Imaging 2005: Image Processing, Pt 1-3, 2005. **5747**: p. 1004-1017.
28. Lange, H. and D.G. Ferris, *Computer-Aided-Diagnosis (CAD) for colposcopy.* Medical Imaging 2005: Image Processing, Pt 1-3, 2005. **5747**: p. 71-84.
29. Li, W., et al., *Using acetowhite opacity index for detecting cervical intraepithelial neoplasia.* J Biomed Opt, 2009. **14**(1): p. 014020.
30. Lotenberg, S., et al., *Automatic evaluation of uterine cervix segmentations - art. no. 65151J.* Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment, 2007. **6515**: p. J5151-J5151.
31. Mehlhorn, G., et al., *Computer-assisted Diagnosis (CAD) in Colposcopy: Evaluation of a Pilot Study.* Anticancer Research, 2012. **32**(12): p. 5221-5226.

32. Mehlhorn, G., et al., *Computer-Assisted Diagnosis in Colposcopy: Results of a Preliminary Experiment?* Acta Cytologica, 2012. **56**(5): p. 554-559.
33. Meyer, F. and A.M.J.V. Kulker, *Automatic Screening of Uterine Cervix Cancer.* Journal De Microscopie Et De Spectroscopie Electroniques, 1987. **12**(1): p. 23-37.
34. Milbourne, A., et al., *Results of a pilot study of multispectral digital colposcopy for the in vivo detection of cervical intraepithelial neoplasia.* Gynecologic Oncology, 2005. **99**(3): p. S67-S75.
35. Pallavi, V. and K. Payal, *Automated analysis of cervix images to grade the severity of cancer.* 2011 Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc), 2011: p. 3439-3442.
36. Park, S.Y., et al., *Automated image analysis of digital colposcopy for the detection of cervical neoplasia.* J Biomed Opt, 2008. **13**(1): p. 014029.
37. Pierangelo, A., et al., *Polarimetric imaging of uterine cervix: a case study.* Opt Express, 2013. **21**(12): p. 14120-30.
38. Srinivasan, Y., et al., *A Unified Model-Based Image Analysis Framework for Automated Detection of Precancerous Lesions in Digitized Uterine Cervix Images.* Ieee Journal of Selected Topics in Signal Processing, 2009. **3**(1): p. 101-111.
39. Srinivasan, Y., et al., *A probabilistic approach to segmentation and classification of neoplasia in uterine cervix images using color and geometric features.* Medical Imaging 2005: Image Processing, Pt 1-3, 2005. **5747**: p. 995-1003.
40. Srinivasan, Y., et al., *Classification of cervix lesions using filter bank-based texture models.* 19th Ieee International Symposium on Computer-Based Medical Systems, Proceedings, 2006: p. 832-837.
41. Tulpule, B., et al., *Segmentation and classification of cervix lesions by pattern and texture analysis.* FUZZ-IEEE 2005: Proceedings of the IEEE International Conference on Fuzzy Systems, 2005: p. 173-176.
42. Zimmerman, G., S. Gordon, and H. Greenspan, *Automatic landmark detection in uterine cervix images for indexing in a content-retrieval system.* 2006 3rd Ieee International Symposium on Biomedical Imaging: Macro to Nano, Vols 1-3, 2006: p. 1348-1351.
43. Zimmerman-Moreno, G. and H. Greenspan, *Automatic detection of specular reflections in uterine cervix images.* Medical Imaging 2006: Image Processing, Pts 1-3, 2006. **6144**.
44. Qiao, L., et al., *Accuracy of visual inspection with acetic acid and with Lugol's iodine for cervical cancer screening: Meta-analysis.* J Obstet Gynaecol Res, 2015. **41**(9): p. 1313-25.
45. Chanthavilay, P., et al., *Accuracy of Combined Visual Inspection with Acetic Acid and Cervical Cytology Testing as a Primary Screening Tool for Cervical Cancer: a Systematic Review and Meta-Analysis.* Asian Pac J Cancer Prev, 2015. **16**(14): p. 5889-97.
46. Jin Huang and C.X. Ling, *Using AUC and accuracy in evaluating learning algorithms.* IEEE Transactions on knowledge and Data Engineering 2005. **17**(3): p. 299-310.
47. Haralick, R.M., Karthikeyan Shanmugam., *Textural features for image classification.* IEEE Transactions on systems, man, and cybernetics, 1973. **6**: p. 610-621.
48. Otsu, N., *Threshold Selection Method from Gray-Level Histograms.* Ieee Transactions on Systems Man and Cybernetics, 1979. **9**(1): p. 62-66.
49. Harris, P.A., et al., *Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support.* J Biomed Inform, 2009. **42**(2): p. 377-81.
50. Das, A., Avijit Kar, and Debasis Bhattacharyya, *Elimination of specular reflection and identification of ROI: The first step in automated detection of Cervical Cancer using Digital Colposcopy.* Imaging Systems and Techniques (IST), 2011 IEEE International Conference on. IEEE, 2011.
51. Jain, A.K., and Farshid Farrokhnia, *Unsupervised texture segmentation using Gabor filters.* Pattern recognition 1991. **24**(12): p. 1167-1186.
52. Arthur, D. and S. Vassilvitskii, *k-means plus plus : The Advantages of Careful Seeding.* Proceedings of the Eighteenth Annual Acm-Siam Symposium on Discrete Algorithms, 2007: p. 1027-1035.
53. Gonzalez, R.C. and R.E. Woods, *Digital image processing*. 3rd ed. 2008, Upper Saddle River, NJ: Pearson/Prentice Hall. xxii, 954 p.
54. Amanpreet Kaur, B.V.K., *Comparison between YCbCr Color Space and CIELab Color Space for Skin Color Segmentation.* International Journal of Applied Information Systems (IJAIS) 2012. **3**(4).

55. *Explanation of the LAB Color Space* [cited 2017 Oct 22]; Available from: http://www.aces.edu/dept/fisheries/education/pond_to_plate/documents/ExplanationoftheLABColorSpace.pdf.
56. [cited 2017 July 24]; Available from: https://www.mathworks.com/help/images/ref/multithresh.html.
57. Suykens, J.A.K., *Least squares support vector machines*. 2002, River Edge, NJ: World Scientific. xiv, 294 p.
58. Kecman, V., *Learning and soft computing : support vector machines, neural networks, and fuzzy logic models*. Complex adaptive systems. 2001, Cambridge, Mass.: MIT Press. xxxii, 541 p.
59. Schölkopf, B. and A.J. Smola, *Learning with kernels : support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. 2002, Cambridge, Mass.: MIT Press. xviii, 626 p.
60. *Selecting Features for Classifying High-dimensional Data*. [cited 2017 July 25]; Available from: https://www.mathworks.com/help/stats/examples/selecting-features-for-classifying-high-dimensional-data.html.
61. Pallavi, V. and K. Payal, *Automated analysis of cervix images to grade the severity of cancer.* Conf Proc IEEE Eng Med Biol Soc, 2011. **2011**: p. 3439-42.
62. Gordon, S., et al., *Content analysis of uterine cervix images: Initial steps towards content based indexing and retrieval of cervigrams - art. no. 61444U.* Medical Imaging 2006: Image Processing, Pts 1-3, 2006. **6144**: p. U1444-U1444.
63. Qureshi, S., V. Das, and F. Zahra, *Evaluation of visual inspection with acetic acid and Lugol's iodine as cervical cancer screening tools in a low-resource setting.* Tropical Doctor, 2010. **40**(1): p. 9-12.
64. Ngoma, T., et al., *Evaluation of cervical visual inspection screening in Dar es Salaam, Tanzania.* Int J Gynaecol Obstet, 2010. **109**(2): p. 100-4.
65. Sankaranarayanan, R., et al., *Test characteristics of visual inspection with 4% acetic acid (VIA) and Lugol's iodine (VILI) in cervical cancer screening in Kerala, India.* Int J Cancer, 2003. **106**(3): p. 404-8.
66. Smith-McCune, K., *Angiogenesis in squamous cell carcinoma in situ and microinvasive carcinoma of the uterine cervix.* Obstet Gynecol, 1997. **89**(3): p. 482-3.
67. Dellas, A., et al., *Angiogenesis in cervical neoplasia: microvessel quantitation in precancerous lesions and invasive carcinomas with clinicopathological correlations.* Gynecol Oncol, 1997. **67**(1): p. 27-33.
68. Hu, F., et al., *Dark field optical imaging reveals vascular changes in an inducible hamster cheek pouch model during carcinogenesis.* Biomed Opt Express, 2016. **7**(9): p. 3247-3261.

Acknowledgements

The authors would like to thank post baccalaureate students, Mark Kellish and Jenna Peters for their invaluable efforts on data collection for this project. The authors would also like to thank Yenny Bellido Fuentes and Catya Lopez, the mid wives at La Liga Contra el Cancer, Lima Peru for their clinical assistance in data collection. This work has been partially supported by the NIH, NSF, and the DOD.

Author contributions

MNA, NR, GS and AS conceived the methods. MNA, CTL, JM worked on hardware and software for the Pocket Colposcope to acquire images used in this study. GV acquired data in Lima Peru. MNA, JM, CTL, GV curated the data. MNA compiled the data and wrote the code for processing and classifying images. MNA, NR, GS, AS and JM analyzed the results. NR and GS acquired the funding for this work. JS and GV served as colposcopy experts in this study. NR, GS and JS supervised this study.

Competing interest statement

NR has founded a company called Athoria and she and other team members (MNA, CTL, JM, MSK) have developed technologies related to this work where the investigators or Duke may benefit financially if this system is sold commercially. NR, MNA, CTL, and JM have filed provisional patents through Duke University, that form the basis for the implementation of the core imaging technology used in the Pocket Colposcope. No algorithms from the manuscript are included in these patent applications. All other authors have declared that no competing interests exist