Cervical Cancer Screening Through a Cloud-Based Algorithmic Analysis of Cervical Images on a Mobile Platform

Satyam[1], Yingying Jiao[1], Yau Ben-Or[2], David Levitz[2], and Roger Zimmermann[1]

1 School of Computing, National University of Singapore, Singapore

2 MobileODT Ltd., Tel Aviv, Israel

Background: For decades, the search for an accurate, low cost, and portable cervical cancer screening test was elusive. The best screening test today - self sampled HPV tests - has higher negative predictive value than positive predictive value, and also suffers from high loss to follow up. Recently, a cervical image pathology classifier based on deep learning was developed to run on a local workstation. However, automated visual evaluation (AVE) algorithms need to run on a mobile platform to make an impact in low resource settings. **Methods**: An AVE classifier was developed using cervical images captured on a mobile colposcope. This classifier is based on the Inception-v3 architecture. It was trained on 577 images, using colposcopic impression by the provider as annotations. The classifier was then validated on another set of 100 images from the same provider. The AVE classifier was deployed on the cloud, called by an application running on a mobile phone. Here, a cervical image captured at the point of care is sent to the cloud, and a binary response is returned. A low precision model was also developed to run natively in the device – this was done by sacrificing accuracy of prediction in favor of faster results without internet access . **Results**: The AVE round trip runtime on the cloud was <1 min, with an AUC of 0.75 and an accuracy of 87%. On the phone the classifier ran in <5 sec, but accuracy dropped to 79%. **Conclusion**: A proof of concept AVE classifier was demonstrated on a mobile platform. Deploying an accurate AVE classifier on a mobile phone has potential to change cervical cancer care.

Introduction

For decades, the search for an accurate, low cost, and portable cervical cancer screening test has been elusive. However, a cervical image pathology classifier based on a deep learning algorithm has the potential to change cervical cancer screening and management. This new test – automated visual evaluation (AVE) – has several appealing advantages, including an accuracy ~20% higher than cytology, nearly instantaneous results, and cost an order of magnitude less than existing tests [1].

AVE, which currently runs on a server, has one simple requirement: a post-acetic acid image of the cervix is needed as input. Thus, to bring AVE to the point of care (PoC), a device is needed that is capable of both capturing a cervical image and running it through a classifier. Mobile phones are ideal for this purpose, as they have both the camera and computational power. Indeed, several mobile colposcopes based on a mobile phone platform already exist. And while there have been several deep learning classifiers running on a mobile platform for consumer applications, an AVE classifier has not been demonstrated on a mobile phone.

However, AVE has distinct challenges that differ from classifiers used in consumer applications:

- High Sensitivity requirement: under-diagnosis results in higher mortality.
- Data Scarcity with High Class Imbalance: Only a small percentage of images are truly positive
- Annotator disagreement: provider opinions vary for both colposcopy and histopathology images
- Variation in Imaging conditions: Even with same device, illumination, glare, and focus all differ
- Deployment at PoC: Deployment on mobile devices introduces a performance tradeoff.

In this work, we demonstrate feasibility of an AVE classifier on a mobile platform. The classifier, based on the Inception-v3 algorithm, was trained on mobile colposcopy images. Image annotations were given by the provider at the PoC. Two implementations were tested on an Android platform: a cloud-based version called by the mobile client, and an algorithm based on TFlite running on the phone itself. This implementation shows promise for deploying AVE algorithm at the PoC.

Methods

Data Collection

- Data source: Images from a retrospective study of routine colpocsopy procedures
- Demographics: conducted at a high-risk population in an urban clinic in California [2].
- Device: the Enhanced Visual Assessment (EVA) System (Fig. 1, from MobileODT).
- Labels: Colposcopic impression (Normal / Cervicitis / Pre-cancerous lesion / Cancer)

Altogether, 3000 images were collected from N=637 patients, which underwent a manual quality review that narrowed it down to 1-2 images per patient. This narrowed the set down to 677 images from 443 patients. Images from 100 patients were separated for validation and the rest were used for training algorithms.

Fig 1: The EVA System's mobile colposcope

Algorithm

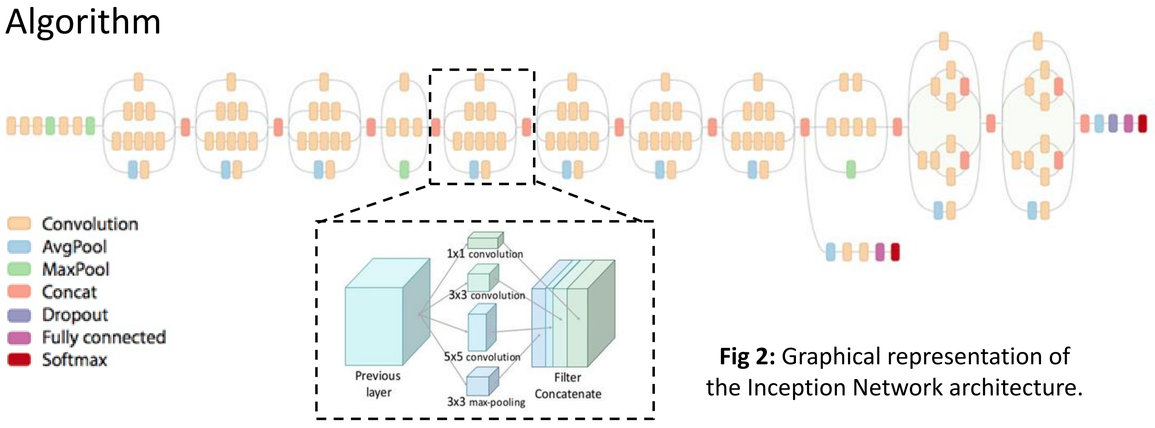

Fig 2: Graphical representation of the Inception Network architecture.

Inception Network parameters

- Number of layers: 311
- Number of inception modules : 9
- Parameter count: 23,885,392 Parameters

Asymmetric Loss Function

The training objective is skewed with a higher penalty for false negatives to account for the need for a high sensitivity algorithm as well as to alleviate the class imbalance problem.

Data Augmentation

- Contrast – channel normalization and contrast limited adaptive histogram equalization (CLAHE) with low clip limit to reduce the appearance of artifacts
- Geometric transformations - Affine transforms (shear, rotation, scale), Perspective transform
- Gamma Correction for low light images

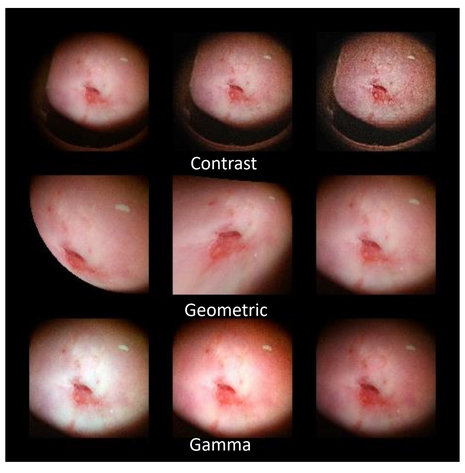

Fig 3: Samples obtained by data augmentation (contrast, geometric and gamma transforms).

Results

Table 1: Performance of cloud-based classifier using raw data and augmented data, as well as on-phone model

	Raw Dataset	Augmented Dataset	On-phone Model
Sensitivity	30%	55%	27.2%
Specificity	76.3%	95%	85.88%
PPV	25%	73.3%	20%
NPV	80.6%	89.4%	90.1%
Accuracy	66.67%	87%	79.16%

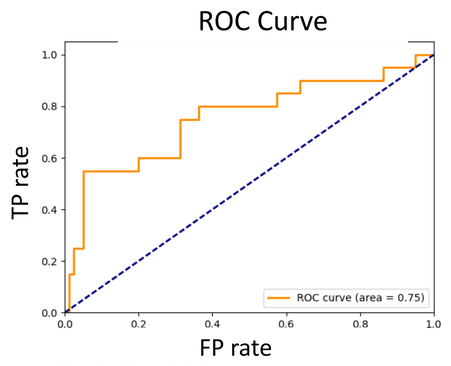

Fig 4: The ROC curve for the best performing network with an AUC of 0.75

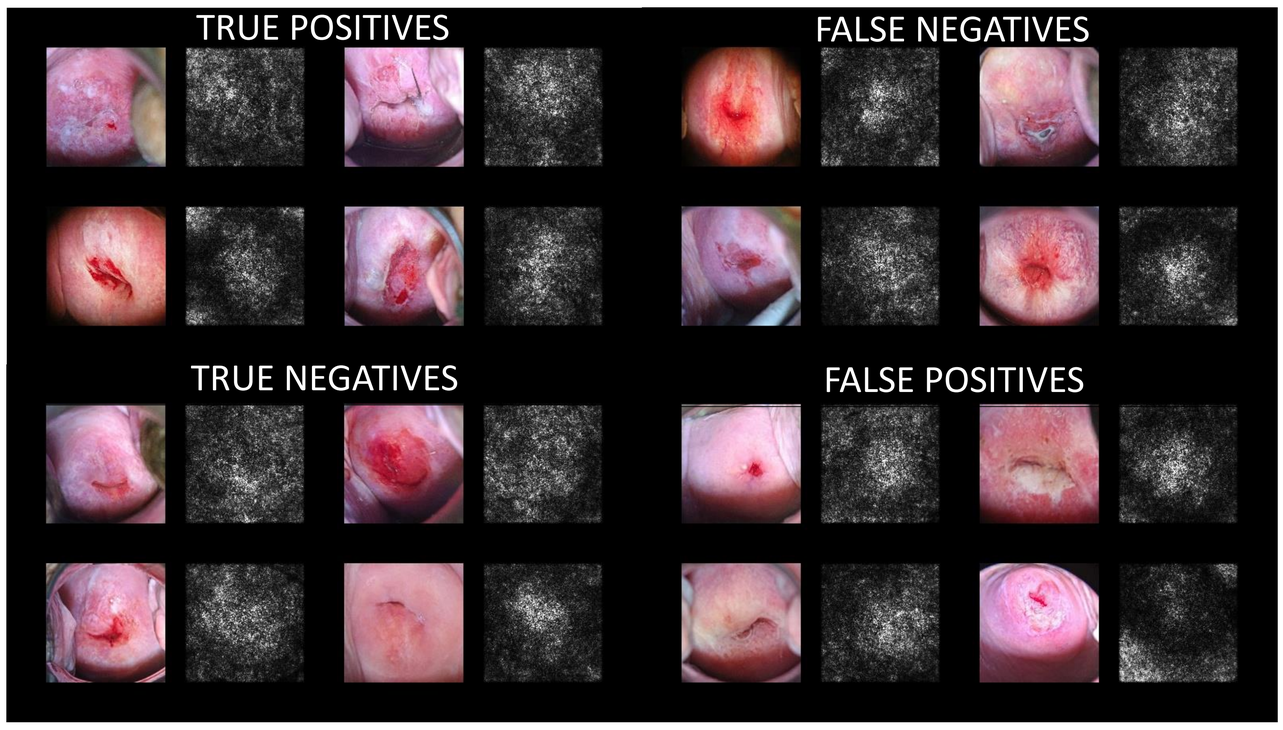

Fig 5: Sample predictions of the Inception network and the corresponding Saliency maps

Discussion and Next Steps

This study demonstrated feasibility of building a classifier of cervical images that runs on a mobile phone platform. The classifier was built around an Inception Network, utilizing images collected by the EVA System during routine colposcopy procedures. On the server, a sensitivity and specificity of 55% and 95% respectively were obtained. In moving the classifier from a server onto a mobile phone, the overall performance was not affected significantly (79.16% vs. 87% accuracy). However, there was a tradeoff in processing time of the classifier – with a latency period of <5 sec for the low-precision On-phone model in comparison to the latency period of >35sec for high-precision computation on a compute cluster in the cloud. This suggests that reasonable classifier performance on can be made available in real time using mobile devices, even when high-speed internet is not available for the cloud computation.

There were 2 main challenges with the current data set. First, the images used to train the classifier came from real colposcopy procedures in which the provider *was not given any instructions* on optimizing image capture in order to train a classifier. We believe such instructions would greatly improve data quality, and boost overall classifier performance. And second, because of the relatively small number of positive (CIN2+) images in the data set, augmentation steps were necessary to build robustness / invariance to the variability in imaging conditions (EVA models, phone models).

Further AVE development will continue in multiple directions. First, histopathological annotations for the images in this study will be used to retrain the classifier, as the lack of ground truth labels for the images was the key limitation of the current study. Second, the classifier will be retrained with a larger pool of images from multiple clinical sites, with different patient demographics. And thirdly, the supervising signal for the training phase will be supplemented with additional information (HPV genotype, E6/E7 oncoproteins, p16 staining). These additional tests will provide a more accurate assessment of the patient's likelihood to develop precancer than histopathology alone, and will ultimately lead to a better performing classifier that can be deployed at the PoC.

Acknowledgements

The authors would like to thank Rachel Steward and Bonnie Betts of FPA Medical for use of their data. The present study was funded by MobileODT.

References

1. Xu T., et al., ``Multi-feature based benchmark for cervical dysplasia classification evaluation,'' In Pattern Recognition, Vol. 63, pp. 468-475, 2017.
2. Steward R., et al. "Comparison of mobile colposcopic impression to cytology and HPV results in a high-risk US population". Presented at IPVC 2018.