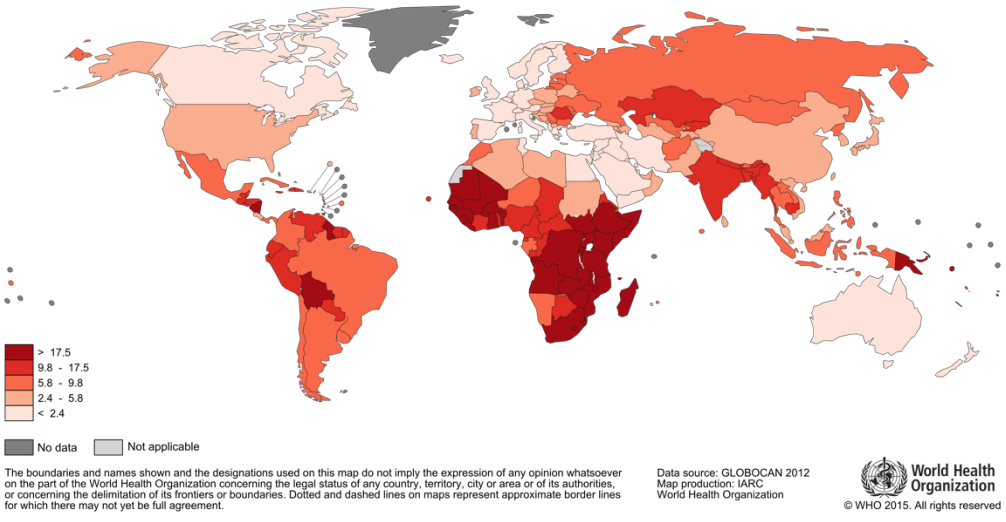
Developing an image-based features recognition algorithm to determine optimal treatment for cervical cancer in low and middle income countries

Keerat Singh

The Academy of Science

Cervical cancer is a disease in which cells along the inner-lining of the lower uterus undergo pre-cancerous changes such as neoplasia and dysplasia (Peirson, Fitzpatrick-Lewis, Ciliska, & Warren, 2013). Cervical cancer is mainly caused by an infectious agent, Human Papilloma Virus (HPV). The two variations, HPV 16 and HPV 18, occur through unprotected sex, and cause over 70% of all cervical cancer cases ("Cervical Cancer," 2016). All sexually-active women have increased chances of obtaining high risk HPV types, thus leading to increased chances for pre-cancerous cell changes. Other risk factors include smoking tobacco, experiencing past history with cancer, and taking the contraceptive pill.

Approximately 500,000 women per year are diagnosed with cervical cancer globally, and 275,000 of those patients die of the disease (Wieringa, Zee, Vries, & Vugt, 2016). The U.S. accounts for a very small percentage of that number (Siegel, Miller, & Jemal, 2015). From 1930 to 2011, cervical cancer mortality rates in the US have decreased over 80% due to medical advances. In low and middle income countries (LMIC), the rates have only reduced 6% from 1930 to 2011. As illustrated on the map below, developing countries such as Kenya, Ethiopia, and Bolivia have much higher mortality rates than developed countries, such as the U.S. and Canada.



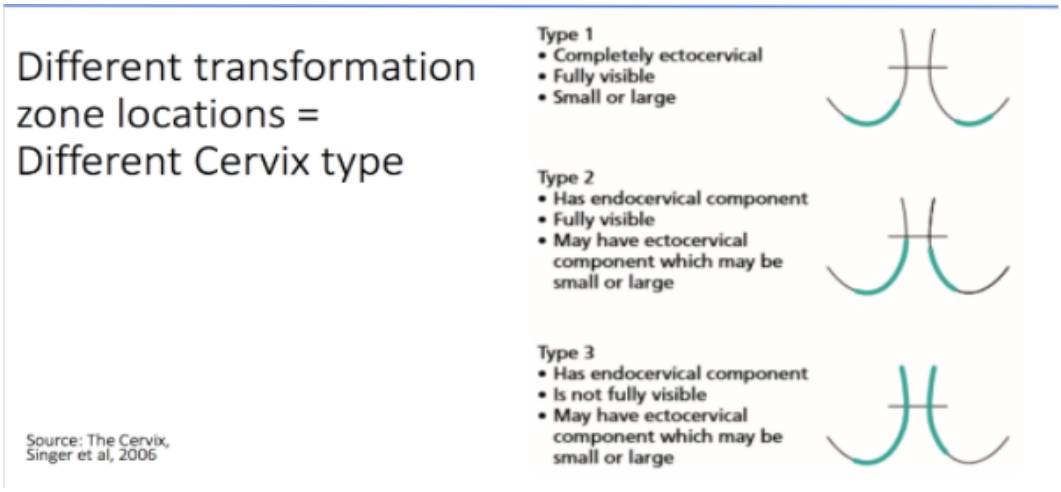
The U.S. has experienced a significant decline in cervical cancer mortality due to advancements in screening and treatment programs (Shulman, 2012). The introduction of Papanicolaou [Pap] diagnostic tests during the twentieth century, combined with the advancement of vaccines, drastically decreased the threat of cervical cancer. Recently, vaccines protecting against the growth of HPV are more expensive, yet more effective, with 90% success rates (Mckee, Bergot, & Leggatt, 2015). The addition of L1 capsid proteins in the vaccine allow neutralizing antibodies to restrict and contain the HPV. When detected early, cervical cancer is one of the easiest diseases to treat, as it only requires a vaccine (Shulman, 2012). Other standard treatments for later stages of the cancer include surgery, radiation therapy, and chemotherapy ("Cervical Cancer Treatment," 2016). The Pap test has allowed doctors in high-resource settings to detect pre-cancerous phenotypes in the cervix before the disease has developed, resulting in very few cases reaching stage I or higher (Mckee, Bergot, & Leggatt, 2015). However, there is a large gap between advanced and LMIC countries for success when preventing and treating cervical cancer (Catarino, 2015). The lack of effective cytology-based screening programs in these low-resource settings is one of the biggest challenges in cervical cancer diagnosis (Saharabuddhe, Parham, Mwanahamuntu, & Vermund, 2011). The main causes of this are the widespread poverty of the countries, and poor training of the practitioners compared to developed countries (Catarino, 2015). Cytology screening requires consistent testing and medical attention that LMIC countries do not yet have the capacity to provide. Furthermore, countries in Eastern Africa and Southern America are also more susceptible to the disease due to unsanitary environments. Additionally, the inability to create an effective and population-wide vaccine allows the HPV to circulate freely (Saharabuddhe, Parham, Mwanahamuntu, & Vermund, 2011). Subsequently, the total number of diagnoses that detect cancer in its earliest

stages are about 60% lower than that of the U.S, leading to the need for more advanced and costly treatments downstream (Catarino, 2015).

To solve this problem, the majority of current research is focused on the analysis and implementation of more efficient screening programs in LMIC countries (Denny et al., 2017). Due to the relatively low medical experience of the practitioners, there have been numerous research projects that have looked into basic but equally-effective diagnosis techniques (Lee, Kang, Ju, 2016). For example, a large topic has been the use of acetic acid for a visual inspection test against the presence of HPV (Denny et al., 2017). Research has also been done with the replacement of acetic acid with Lugol's iodine (Vyas, Bhalodia, & Thakor, 2016). Both visual inspection techniques have had much more success in middle-income countries, such as parts of India and Pakistan. A large percentage of screening programs that are currently being looked into for LMIC countries aren't feasible; despite being reliable in terms of accuracy, they require complex human and financial infrastructures to be implemented and maintained (Lee, Kang, Ju, 2016). In Sub-Saharan Africa alone, it was reported in 2016 that less than 5% of at-risk women had been screened, indicating new screening programs' lack of implementation and success (Catarino, 2015). As a result, most cases of cervical cancer become invasive to the point that they are unable to treat. Many women in these parts of the world are receiving treatments that are ineffective and many times harmful to them ("Intel & MobileODT," 2017). Health providers often do not have the necessary skills, and more importantly, diagnostic resources, to recognize which treatments to use to prevent the spread of the cancer. The inability to discern which treatment will work for a woman based on her type of cervix has led to increased mortality, unnecessary expenses and the wastage of materials. False treatments and the lack of

effective screening programs are substantial reason why most cases of cervical cancer in LMIC countries reach stage III and above, usually resulting in death (Denny et al., 2017).

While most current research involves modifying the screening process to catch the disease early, relatively little is concerned with the advanced treatment process itself. Currently, there does not yet exist a cheap and accurate technology to determine a woman's cervix type, thus allowing physicians to decide on the appropriate treatment ("Intel & MobileODT," 2017). Creating this type of product will limit ineffective and costly treatments, while also improving the cervical cancer mortality rate in LMIC countries. Therefore, my goal is to create and train a features identification algorithm that will accurately identify a woman's cervix type (1, 2, or 3) based on a given image. The three cervix types are indicated below.



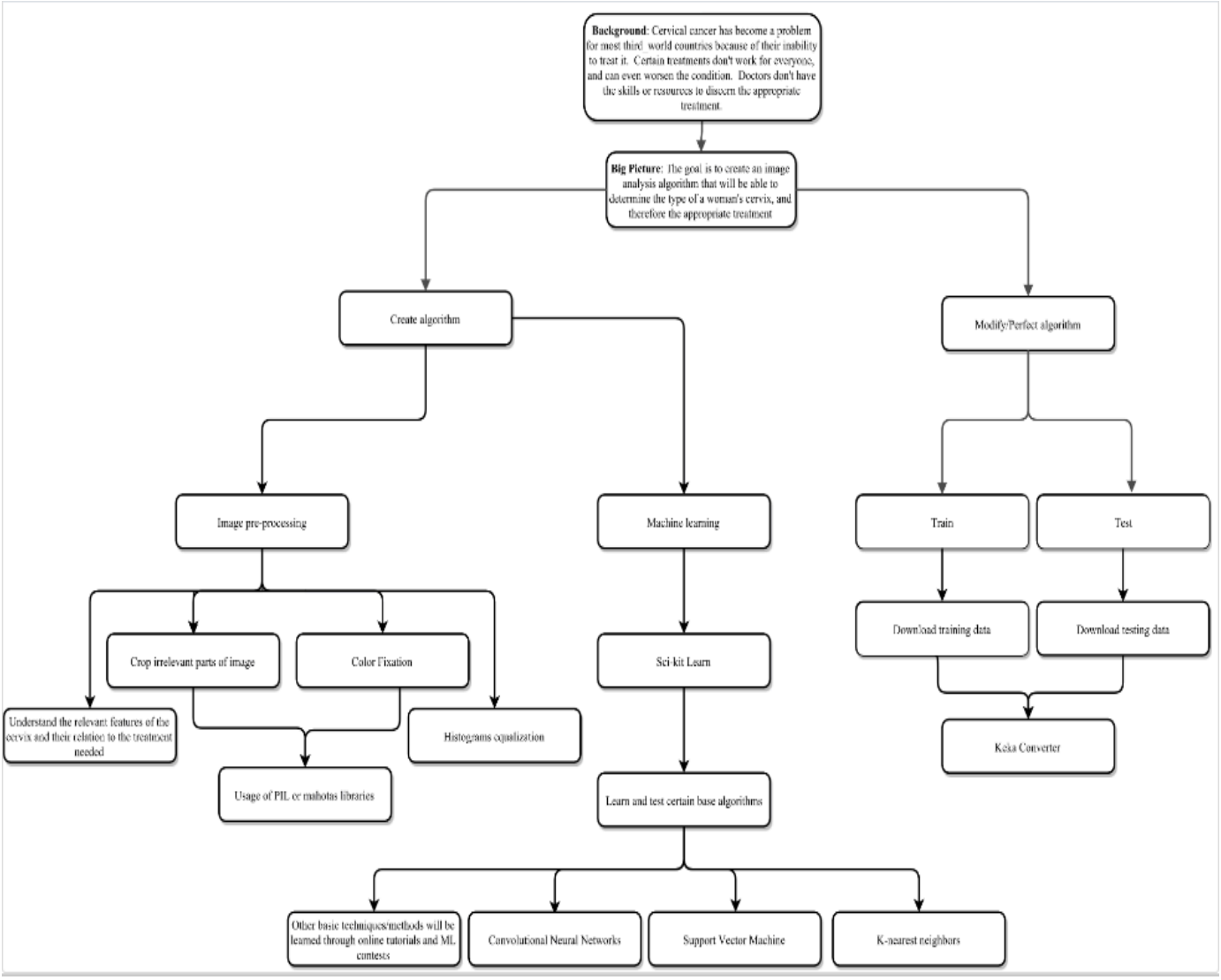
The imaging technology will be able to identify the type of cervix based on its transformation zone location, resulting in higher success prescribing treatment. In essence, all the image classifier will do is it will analyze the numerical properties of different categorical image features and determine what matches the pattern for each type. What is being analyzed and how the analysis occurs are what differentiate solutions from each other. The algorithm will mainly factor in the location and shape of the transformation zone and whether it is fully visible or not.

The main problem that comes with this data is that there are not many distinguishing factors between the three types. This will hopefully be accommodated for with the large size of the training and testing datasets.

All of the images that will be used to train the algorithm are coming from the Kaggle database. The data files are available through .txt and .7z files, meaning that I will need to download a .7z file converter to access parts of it. The cervices within the images are all considered non-cancerous, but they all showcase the individual transformation zone locations, which ultimately determine the different cervix types. Once the data is fully accessible, I will need to create the first part of the software that will process the images. I have extensive experience with the Python language, but I will need to learn how to use certain libraries that help in analyzing images. These libraries primarily include OpenCV, NumPy, SciPy, and Pillow. The main library, OpenCV, has a full online walkthrough of all of the library's features that I plan on working through before proceeding. This tutorial contains OpenCV lessons on the image processing functions, image input and output, the 2D features framework, deep neural networks, and the machine learning classes of the library. While the tutorial does contain brief introductions to the deep learning aspects of it, it will mainly serve as an introduction to the library itself. To learn the image classification skills needed to create the algorithm, I will take the online CV-Tricks tutorials that explain how to use different pre-existing algorithms, such as convolutional neural networks, a support vector machine, and K nearest neighbors. In addition, Kaggle has multiple machine learning sample projects that serve as useful introductions. I feel that it is important to get a basic understanding of the main part of my project before delving into the other subsections of it. Therefore, the first thing I will do is going to be to complete two of the introductory problems offered by Kaggle (Titanic and Facebook Recruiting). By learning the

basics of machine learning before starting, I will have a better understanding of how the initial image pre-processing will fit in with the rest of the project.

Below is a diagram of each objective moving forward in the project. It is essentially a map of everything mentioned in the above paragraph with some minor additions.



My end goal with this project is to create a fully-functioning, easily accessible application that will take in an image of a cervix and output its type based on its transformation zone. Finishing the project with an app is the optimal scenario because that is how my research could most help

those countries in need. I have pre-existing experience with Java and Swift, which are the languages I plan to use if Python isn't sufficient. This cheap and reliable technology can become an invaluable asset in the cervical cancer toolkit for healthcare providers in less developed regions of the world.

References

Catarino, R. (2015). Cervical cancer screening in developing countries at a crossroad: Emerging technologies and policy choices. *World Journal of Clinical Oncology,6*(6), 281. doi:10.5306/wjco.v6.i6.281

Cervical cancer. (2016). Retrieved May 18, 2017, from http://www.cancerresearchuk.org/aboutcancer/cervical-cancer/risks-causes

Cervical Cancer Treatment. (2016). Retrieved May 18, 2017, from https://www.cancer.gov/types/cervical/patient/cervical-treatment-pdq#section/_180

Denny, L., Sanjose, S. D., Mutebi, M., Anderson, B. O., Kim, J., Jeronimo, J., . . . Sankaranarayanan, R. (2017). Interventions to close the divide for women with breast and cervical cancer between low-income and middle-income countries and high-income countries. *The Lancet,389*(10071), 861-870. doi:10.1016/s0140-6736(16)31795-0

Intel & MobileODT Cervical Cancer Screening. (2017.). Retrieved May 18, 2017, from https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening

Lee, H., Kang, Y., & Ju, W. (2016). Cervical Cancer Screening in Developing Countries: Using Visual Inspection Methods. *Clinical Journal of Oncology Nursing,20*(1), 79-83. doi:10.1188/16.cjon.79-83

Mckee, S. J., Bergot, A., & Leggatt, G. R. (2015). Recent progress in vaccination against human papillomavirus-mediated cervical cancer. *Reviews in Medical Virology,25*, 54-71. doi:10.1002/rmv.1824

Peirson, L., Fitzpatrick-Lewis, D., Ciliska, D., & Warren, R. (2013). Screening for cervical cancer: a systematic review and meta-analysis. *Systematic Reviews,2*(1). doi:10.1186/2046-4053-2-35

Sahasrabuddhe, V. V., Parham, G. P., Mwanahamuntu, M. H., & Vermund, S. H. (2011). Cervical Cancer Prevention in Low- and Middle-Income Countries: Feasible, Affordable, Essential. *Cancer Prevention Research,5*(1), 11-17. doi:10.1158/1940-6207.capr-11-0540

Shulman, L. (2012). American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology Screening Guidelines for the Prevention and Early Detection of Cervical Cancer. *Yearbook of Obstetrics, Gynecology and Women's Health,2012*, 339-342. doi:10.1016/j.yobg.2012.06.159

Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians,65*(1), 5-29. doi:10.3322/caac.21254

Vyas, K., Bhalodia, K., & Thakor, N. (2016). Role of Pap's smear and visual inspection of cervix with Lugol's iodine for early detection of premalignant and malignant lesions of cervix: a cross sectional study. *International Journal of Reproduction, Contraception, Obstetrics and Gynecology,*2684-2686. doi:10.18203/2320-1770.ijrcog20162646

Wieringa, H. W., Zee, A. G., Vries, E. G., & Vugt, M. A. (2016). Breaking the DNA damage response to improve cervical cancer treatment. *Cancer Treatment Reviews,42*, 30-40. doi:10.1016/j.ctrv.2015.11.008