

Cervical Cancer

Sean Wade

A project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Science

David Wingate, Chair
TODO
TODO

Department of Computer Science
Brigham Young University

Copyright © 2019 Sean Wade
All Rights Reserved

ABSTRACT

Cervical Cancer

Sean Wade

Department of Computer Science, BYU

Master of Science

Cervical cancer is one of the deadliest cancers among women worldwide. Manual screening is relatively straightforward, however it is time-consuming and expensive. The goal of this project is to help develop a low cost sensor and algorithm to detect this cancer.

Keywords: segmentation, CNN, synthetic data, data augmentation, cervical cancer, medical imaging

Table of Contents

1	Introduction	1
1.1	The pap-smear screening	1
1.2	Cervix Stuff	1
2	Datasets	2
2.1	SIPaKMeD	2
2.2	Herlev	2
3	Background	4
4	Data Generation	5
5	Experiments and Reuslts	6
6	Future Work	7
	References	8

Chapter 1

Introduction

Cervical cancer accounts for 6.6% of cancer cases in the world. In 2018 alone, there were well over 570,000 cases. Even worse is the disproportionate way populations are effected. About 90% of deaths from cervical cancer came from low to middle income countries.[1]

Fortunately the high mortality rate from cervical cancer can be reduced by effective screening and early treatment.

1.1 The pap-smear screening

The pap-smear screening was developed by Georges Papanicolaou. The test is done by collecting cells and staining them. When they are looked at under a microscope cytotechnicians are able to detect pre-cancerous cells in the uterine cervix. Human labeling and classification of the individual cells is very time consuming and expensive since it must be performed by skilled cytotechnicians. The goal of this project is to not do that.

Using a small brush, cotton-stick or wooden stick, a cytological sample is taken from the cervix and smeared onto a thin glass slide. To clarify the cells characteristics the smear is stained using the Papanicolaou method, so the different components of the cells are emphasized with specific colors - this makes it more clear in a microscope.[4]

Each microscope slide contains up to 300,000 single cells with different orientations and overlap[4]

1.2 Cervix Stuff

Chapter 2

Datasets

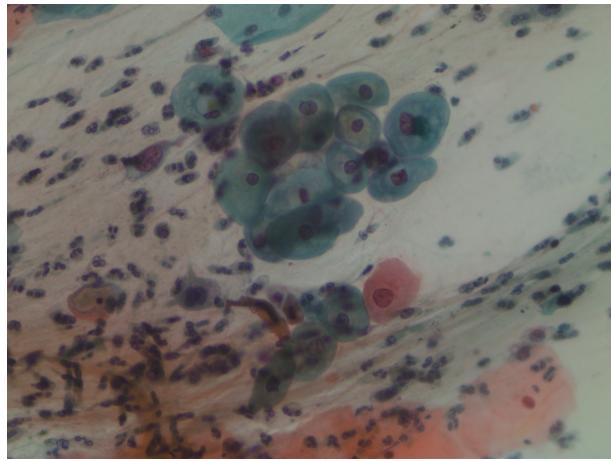
The datasets used are constructed by cyto-technicians for classification purposes. These technicians use a microscope with a resolution of $0.201 \mu\text{m} / \text{pixel}$ to grab every cell.

2.1 SIPaKMeD

The SIPaKMeD dataset consists of 4049 single cell images which have been manually cropped from 996 cluster cell images of Pap smear slides. These cells are then labeled in 5 categories: (1) Dyskeratotic, (2) Koilocytotic, (3) Metaplastic, (4) Parabasal and (5) Superficial-Intermediate. 1-2 classes represent the abnormal cervical cells, 4-5 classes represent normal cervical cells, and 3 represents the benign cells.[5]

2.2 Herlev

The Herlev dataset is comprised of 917 isolated single cell images. These are distributed unequally between seven classes of cells. (1) Superficial squamous epithelia, (2) Intermediate squamous epithelia, (3) Columnar epithelial, (4) Mild squamous non-keratinizing dysplasia, (5) Moderate squamous non-keratinizing dysplasia, (6) Severe squamous non-keratinizing dysplasia and (7) Squamous cell carcinoma in situ intermediate. The classes 1-3 are normal cervical cells whereas classes 4-7 are abnormal cervical cells.[2]



Chapter 3

Background

Current automatic screening systems that incorporate pap smear have a very similar workflow: cell segmentation, cytoplasm and nuclei segmentation, feature extraction, and cell classification.

Whole slide classification (no segmentation) [3]

Chapter 4

Data Generation

- rotations
- flip horizontal and vertical
- random crop and zoom

Chapter 5

Experiments and Reuslts

Chapter 6

Future Work

References

- [1] World health organization. cervical cancer, 2018. [online; accessed 12-december-2018].
- [2] Jan Jantzen, Jonas Norup, Georgios Dounias, and Beth Bjerregaard. Pap-smear benchmark data for pattern classification. In *Proc. NiSIS 2005*, pages 1–9. NiSIS, 2005.
- [3] Kranthi Kiran GV and G Meghana Reddy. Automatic classification of whole slide pap smear images using cnn with pca based feature interpretation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
- [4] Jonas Norup. Classification of pap-smear data by tranduction neuro-fuzzy methods. 2005.
- [5] Marina Plissiti, P. Dimitrakopoulos, Giorgos Sfikas, Christophoros Nikou, O. Krikoni, and Antonia Charchanti. Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. pages 3144–3148, 10 2018. doi: 10.1109/ICIP.2018.8451588.