CERVIX CANCER DETECTION BASED ON PATTERN RECOGNITION IN CERVICAL CYTOLOGICAL SLIDE IMAGES

By

Jeremiah Suryatenggara

A Bachelor's Thesis
Submitted to the Faculty of

LIFE SCIENCE

in partial fulfillment of the requirements for the Degree of

BACHELOR OF SCIENCES
WITH A MAJOR IN BIOMEDICAL ENGINEERING

SWISS GERMAN UNIVERSITY

SWISS GERMAN UNIVERSITY
Campus German Centre
Bumi Serpong Damai – 15321
Island of Java, Indonesia
www.sgu.ac.id

July 2009

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.

Jeremiah Suryatenggara Date

Approved by:

Dr. Bernadetta Kwintiana Ane, Advisor Date

SWISS GERMAN UNIVERSITY

Dr. rer. nat. Maruli Pandjaitan, Co-Advisor Date

Chairman of the Examination Steering Committee Date

ACKNOWLEDGMENTS

First, I would like to thank God for His blessings, guidance and protection upon me throughout the whole research work and thesis preparation.

In this opportunity, I would also like to thank Dr. rer. nat. Maruli Pandjaitan, the Dean of Life Science in SGU and also the co-advisor of this thesis work, and Dr. Bernadetta Kwintiana Ane, the advisor of this thesis work, for their ideas and inputs that played a big part in the completion of this thesis.

I would also like to give a special thanks to my friend, Henry Kho the Great, for his great generosity and great programming skill has been a great help for the completion of this thesis.

Thank you to my family for their comfort and endless support. Lots of thank is also addressed to all my friends in SGU for the wonderful memories for the last four years and their supports during my study at SGU.

Last but not least, I would like to thank SGU for the opportunity to work on the research of cervical cancer detection.

SWISS GERMAN UNIVERSITY

ABSTRACT

By

Jeremiah Suryatenggara

SWISS GERMAN UNIVERSITY
Bumi Serpong Damai

Dr. Bernadetta Kwintiana Ane, Advisor
Dr. rer. nat. Maruli Pandjaitan, Co-Advisor
Dr. Winfried Steinberg, Co-Advisor

Cervical cancer is the most common type of cancer in women throughout the world. This cancer is largely preventable through screening during its early stages and appropriate treatment. The effective screening test for detecting pre-malignant cells in uterine cervix before the progress into cancer is called Pap smear.

Pap smear using Papanicolau staining method emphasizes the different components of the cells with specific colors. Available cell features like morphology, shape and color of nucleus and cytoplasm are used for the classification of Pap smear image into normal and abnormal cells.

However, the screening which is done manually by the cytologist was inefficient and time consuming. It requires high skill and experiences of the cytologists. This requirement reads to a condition that the diagnosis inherently prone to the human error. Coping with this problem, this research is to introduce an automated diagnosis

algorithm for early detection of cervix cancer. The diagnosis algorithm is developed to recognize pattern on 2-dimentional digital cervical cytological image produced from Pap smear test. Pattern recognition is applied to variables of cell morphology and color intensity. Afterwards, measurement and identification of cells into normal and abnormal class is done on the basis of parameters color intensity, N/C ratio, and 2D wavelet approximation coefficients.

The automated diagnosis algorithm is intended to improve reliability as well as reduce time consumption in the conventional diagnosis of cervix cancer. Therefore, this will produce more accurate, faster and less expensive analysis of Pap smear test, which can be used to provide better health service for public community.

SGU®
SWISS GERMAN UNIVERSITY

TABLE OF CONTENTS

	Page
Cover page	1
Statement by the Author	2
Acknowledgement	3
Abstract	4
Table of Contents	6
List of Figures	11
List of Tables	14
CHAPTER 1 – INTRODUCTION	**15**
1.1 Background	15
1.2 Research Problems	17
1.3 Research Objective	17
1.4 Benefits of Research	17
1.4.1 Public community	17
1.4.2 Medical community	18
1.4.3 Academic community	18
1.5 Research Assumption and Constraints	18
1.5.1 Assumption	18
1.5.2 Constraints	18
1.6 Methodology	18
1.6.1 Literature study	18
1.6.2 Image processing	18
1.6.3 Feature extraction	19
1.6.4 Pattern recognition	19
1.6.5 Classification	19
1.6.6 Result visualization	19
1.6.7 Process automation	19
1.7. Organization of Thesis	19
1.7.1. Chapter 1 – Introduction	19
1.7.2. Chapter 2 – Literature Study	20
1.7.3. Chapter 3 – Methodology	20

SWISS GERMAN UNIVERSITY

Page

1.7.4. Chapter 4 – Research Findings and Analysis 20
1.7.5. Chapter 5 – Conclusion and Recommendation 20

CHAPTER 2 - LITERATURE STUDY 21
2.1 Cervical Cancer 21
2.2 Pap Smear Test 23
2.3 Cervical Cells 24
2.4 Image Acquisition 25
2.4.1 Image acquisition using sensor arrays 25
2.4.2 Acquiring standardized images 26
2.4.3 Capturing digitized images from video signal 26
2.5 Image Processing 26
2.5.1 Image enhancement 27
2.5.1.1 Contrast stretching 28
2.5.1.2 Histogram processing 29
2.5.2 Image filtering 29
2.5.3 Image segmentation 30
2.5.2.1 Edge detection 31
2.5.2.2 Thresholding 32
2.6. Feature Extraction 33
2.7. Pattern Recognition 33
2.8. Wavelet 34
2.8.1 Wavelet transform 35
2.8.2 Wavelet families 36
2.9 Classification 36
2.9.1 k-Nearest Neighbor (k-NN) classifier algorithm 37
2.9.2 Support Vector Machine (SVM) classifier algorithm 40
2.9.3 Naïve Bayes (NB) classifier algorithm 43

CHAPTER 3 – METHODOLOGY 46
3.1 Image Acquisition 47
3.1.1 Preparation of cervix cells specimen 48
3.1.2 Preparation of digital imaging devices 48

SWISS GERMAN UNIVERSITY

Page

3.1.3 Acquisition of digital cervix cell images 50
3.1.3.1 Specimen visualization 50
3.1.3.2 Digital cervix cell image acquisition 50
3.1.3.3 Digital cervix cells image archiving 51
3.2 Image Processing 51
3.2.1 Image retrieval and conversion 52
3.2.2 Image enhancement 52
3.2.3 Image segmentation 55
3.2.4 Image filtering 57
3.2.4.1 Image erosion 58
3.2.4.2 Image reconstruction 59
3.3 Feature Extraction 59
3.3.1 Determination of the required features 59
3.3.2 Extraction of the required features 60
3.4 Pattern Recognition 62
3.4.1 Observation and analysis of unique patterns 62
3.4.2 Comparison and confirmation of the patterns 65
3.4.3 Removal of all misleading factors in the patterns 68
3.4.4 Conversion of patterns into classifiable variables 70
3.4.5 Labeling of the attributes 73
3.4.6 Establishment of training database 74
3.5 Classification 74
3.5.1 Selection of appropriate classification algorithm 75
3.5.2 Setting of chosen classification algorithms 78
3.5.3 Cross validation of the classification results 79
3.5.4 Classification of the observed cells into groups 80
3.5.4.1 k-Nearest Neighbor classification algorithm 81
3.5.4.2 Support Vector Machine classification algorithm 81
3.5.4.3 Naïve Bayes classification algorithm 82
3.6 Result Visualization 83
3.6.1 Visualization of diagnostic images 83
3.6.2 Visualization of diagnostic graphs and histograms 84
3.6.3 Visualization of diagnostic pattern values 85

SWISS GERMAN UNIVERSITY

Page

3.6.4 Visualization of observed cell's classification results 85
3.7 Process Automation 86
3.7.1 Formulation of the workflow of the whole proces 86
3.7.2 Encoding of the system with programming language 86
3.7.3 Construction of a user interactive GUI 87
3.7.4 Test-running of the prototyped software 88

CHAPTER 4 – RESEARCH FINDINGS AND ANALYSIS 89
4.1 Findings in Pattern Recognition 89
4.1.1 Pattern recognition based on color intensity 89
4.1.2 Pattern recognition based on N/C ratio 91
4.1.3 Pattern recognition based on wavelet characterisitics 93
4.1.4 Pattern values trend of normal cells 96
4.1.4.1 Normal cell color intensity pattern values 97
4.1.4.2 Normal cell N/C ratio pattern values 97
4.1.4.3 Normal cell wavelet characteristics pattern values 98
4.1.5 Pattern values trend of cancerous cells 99
4.1.5.1 Cancerous cell color intensity pattern values 99
4.1.5.2 Cancerous cell N/C ratio pattern values 100
4.1.5.3 Cancerous cell wavelet characteristics pattern values 100
4.2 Findings in Classification 101
4.2.1 Training database cross validation benchmarking results 101

CHAPTER 5 – CONCLUSION AND RECOMMENDATION 104
5.1 Conclusion 104
5.1.1 Attributes of normal cells 104
5.1.1.1 Based on conventional procedure 104
5.1.1.2 Based on automated procedure 104
5.1.2 Attributes of cancerous cells 105
5.1.2.1 Based on conventional procedure 105
5.1.2.2 Based on automated procedure 105
5.1.3 Advantages and disadvantages of conventional procedure 106
5.1.3.1 Advantages 106

SWISS GERMAN UNIVERSITY

Page

5.1.3.2 Disadvantages 106
5.1.4 Advantages and disadvantages of automated procedure 107
5.1.4.1 Advantages 107
5.1.4.2 Disadvantages 108
5.2 Recommendation 108

GLOSSARY 109

REFERENCES 111

APPENDIX 113

CURRICULUM VITAE 139

SGU ®
SWISS GERMAN UNIVERSITY