

# METODE DATA MINING UNTUK KLASIFIKASI DATA SEL NUKLEUS DAN SEL RADANG BERDASARKAN ANALISA TEKSTUR

**Toni Arifin**

Universitas BSI Bandung, toni.tfn@bsi.ac.id

**ABSTRACT** - The Pap Smear test is done to see the presence of infection or changes in cells that can turn into cancer cells. In this research is using data on analysis results of texture image processing on previous research that is using a nucleus cell and inflammation cell in the image Pap Smear cell. The purpose of this research is to find the best method for classifying the nucleus cell and inflammation cell based on texture analysis GLCM (Gray Level Co-occurrence Matrix) in this research used of method Decision tree (C4.5), Naive Bayes and k-Nearest Neighbour. The results of this research brings about the best methods for classification of the data nucleus cell and inflammation cell that is a method of Decision tree (C4.5) with accuracy 97,56% whereas results for Naive Bayes 90,89% and k-Nearest Neighbour 95,97%.

**Keywords:** Data mining, classification, Pap Smear cell, Texture Analysis

**ABSTRAKSI** - Tes *Pap Smear* dilakukan untuk melihat adanya infeksi atau perubahan sel-sel yang dapat berubah menjadi sel kanker. Pada penelitian ini menggunakan data analisis tekstur yang didapatkan dari hasil pengolahan citra pada penelitian sebelumnya yaitu menggunakan sel nukleus dan sel radang pada citra sel *Pap Smear*. Tujuan dari penelitian ini adalah mencari metode terbaik untuk mengklasifikasikan sel nukleus dan sel radang berdasarkan analisa teksur GLCM (Gray Level Co-occurrence Matrix) Metode yang digunakan dalam penelitian ini adalah metode *Decision tree* (C4.5), *Naive Bayes* dan *k-Nearest Neighbour*. Hasil dari penelitian ini didapatkan metode terbaik untuk klasifikasi data sel nukleus dan sel radang yaitu metode *Decision tree* (C4.5) dengan akurasi 97,56% sedangkan hasil untuk *Naive Bayes* 90,89% dan *k-Nearest Neighbour* 95,97%.

**Kata Kunci:** Data mining, Klasifikasi, Sel *Pap Smear*, Analisa Tekstur

## I. PENDAHULUAN

Kanker serviks merupakan salah satu penyakit berbahaya di dunia dan merupakan isu utama mengenai kesehatan kaum wanita, terutama di negara berkembang, dibutuhkan bertahun-tahun untuk sel abnormal untuk tumbuh, dan setiap wanita perlu tahu tanda awal kelainan sel sebagai pencegahan awal. Virus HPV (Human Papilloma Virus) adalah penyebab utama kanker serviks. Virus HPV memiliki lebih dari 100 tipe virus di mana sebagian besar diantaranya tidak berbahaya dan akan lenyap dengan sendirinya, jenis virus HPV yang menyebabkan kanker serviks dan yang paling fatal adalah virus HPV tipe 16

dan 18. Selain disebabkan oleh virus HPV, sel-sel abnormal pada leher rahim juga bisa tumbuh akibat paparan radiasi atau pencemaran bahan kimia yang terjadi dalam jangka waktu cukup lama. Melihat begitu tingginya jumlah wanita yang terkena kanker serviks ada beberapa cara dalam mendeteksi secara dini kanker serviks salah satunya adalah melakukan pemeriksaan *Pap Smear*. Pengamatan sel *Pap Smear* secara manual akan memakan waktu lama, melelahkan dan rawan kesalahan itu semua salah satunya diakibatkan karena adanya sel radang yang mengganggu dalam pengamatan terutama pengamatan pada sel nukleus. Penelitian ini dilakukan untuk mengetahui metode

data mining mana yang terbaik dalam mengklasifikasikan data sel nukleus dan sel radang berdasarkan analisis tekstur *Gray Level Co-occurrence Matrix*. Data analisa tekstur pada penelitian ini diambil dari penelitian sebelumnya yang dilakukan oleh (Arifin et. al, 2015).

## I. KAJIAN LITERATUR

Kanker serviks dimulai dalam sel pada permukaan serviks atau leher rahim. Dengan berjalannya waktu, kanker serviks dapat menyerang lebih jauh ke dalam serviks dan jaringan didekatnya. Pada umumnya kanker serviks berkembang dari sebuah kondisi pra-kanker. Pra kanker ini timbul ketika serviks terinfeksi oleh HPV (*Human Pappilloma Virus*). Kebanyakan pra-kanker lenyap dengan sendirinya, tetapi jika sel pra kanker bertahan dan tidak diobati, dapat menjadi kanker (Ahmed redhwan, 2012). Sudah banyak penelitian yang telah dilakukan untuk klasifikasi sel nukleus dan sel radang diantaranya adalah penelitian yang menggunakan pendekatan GLCM (*Gray Level Co-occurrence Matrix*). Analisis tekstur difokuskan pada citra inti sel *Pap Smear*, dengan menggunakan Matlab. Ada 5 parameter yang digunakan yaitu: kontras, korelasi, energi, homogenitas dan entropi. (Pratama et al, 2012), Selanjutnya penelitian tersebut dilanjutkan kembali untuk klasifikasi berdasarkan analisa tekstur menggunakan *Decision tree* dan penambahan atribut yang didapat dari perhitungan *brightness*. (Arifin et al, 2013). Dan penelitian selanjutnya adalah metode analisa tekstur untuk klasifikasi sel nukleus dan sel radang pada citra mikroskopik pap smear (Arifin et al, 2015). Pada penelitiannya menggunakan metode analisa tekstur *Gray Level Co-occurrence Matrix*.

### A. Data Mining

Kegiatan data mining merupakan proses yang berulang-ulang yang ditujukan untuk analisis database yang besar, dengan tujuan untuk menggali informasi dan pengetahuan yang akurat dan berpotensi untuk pengambilan keputusan dan pemecahan masalah (Vercellis, 2009). Data mining

didefinisikan sebagai proses menemukan dan menggambarkan pola struktural dalam data sebagai alat untuk membantu menjelaskan data dan membuat prediksi dari data tersebut (Witten & Frank, 2005). Data mining merupakan bagian dari Knowledge Discovery Data (KDD) yang merupakan proses ekstraksi informasi yang berguna, tidak diketahui sebelumnya, dan tersembunyi dari data (Bramer, 2013) dan juga mengembangkan model yang digunakan untuk memahami fenomena dari analisis data dan prediksi (Maimon & Rokach, 2010). Pada penelitian ini akan dilakukan perbandingan untuk mengetahui metode data mining mana yang terbaik untuk mengklasifikasikan data sel nukleus dan sel radang pada citra sel *Pap Smear*. Metode data mining yang digunakan diantaranya sebagai berikut:

#### 1. *Decision tree* (C4.5)

Algoritma ini diperkenalkan oleh J. Ross Quinlan yang merupakan pengembangan dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar, node akar tidak memiliki inputan. Node lain yang bukan sebagai akar tetapi memiliki tepat satu inputan disebut node internal atau test node, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu class (Maimon et al., 2010). Langkah-langkah membangun pohon keputusan menggunakan algoritma C4.5. Pemilihan atribut sebagai akar berdasarkan pada nilai *gain* tertinggi dari atribut yang ada. Untuk menghitung nilai *gain* tertinggi digunakan persamaan berikut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Dimana:

S : Himpunan kasus  
A : Atribut

n : Jumlah partisi atribut A  
 |Si| : Jumlah kasus pada partisi ke-i  
 |S| : Jumlah kasus dalam S

Selanjutnya adalah rumus mencari nilai entropi:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Dimana:

S : Himpunan kasus  
 n : Jumlah partisi S  
 Pi : Proporsi dari Si terhadap S

### 2. Naive Bayes

*Naive Bayes* merupakan metode yang tidak memiliki aturan (Alfisahrin, 2014). *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training. *Naive Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam data mining (Wu & Kumar, 2009). Bentuk umum dari teorema bayes dapat dilihat pada persamaan berikut:

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \quad (3)$$

Dimana:

X: Data dengan class yang belum diketahui  
 H: Hipotesis data X merupakan suatu class spesifik  
 P(H|X): Probabilitas hipotesis H berdasarkan kondisi X (posterior probability)  
 P(H): Probabilitas hipotesis H (prior probability)  
 P(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H  
 P(X): Probabilitas dari X

### 3. k-Nearest Neighbour

Algoritma *k-Nearest Neighbour* adalah suatu metode yang menggunakan algoritma supervised (Wu & Kumar, 2009). Tujuan dari algoritma *k-Nearest Neighbour* adalah untuk mengklasifikasi objek baru berdasarkan atribut dan

training samples (Larose, 2005). Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *k-Nearest Neighbour*. Pada proses pengklasifikasian, algoritma ini tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Algoritma *k-Nearest Neighbour* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru.

*Euclidean Distance*

$$d(x, y) = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2} \quad (4)$$

*Manhattan Distance*

$$d(x, y) = \sqrt{\sum_{k=1}^n |X_k - Y_k|} \quad (5)$$

Untuk mengukur jarak dari atribut yang mempunyai nilai besar, seperti atribut pendapatan, maka dilakukan normalisasi. Normalisasi bisa dilakukan dengan *min-max normalization* atau *Z-score standardization* (Larose, 2005). Jika data training terdiri dari atribut campuran antara numerik dan kategori, lebih baik gunakan *min-max normalization* (Larose, 2005).

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (6)$$

$$x^* = \text{Standarisasi min-max} \quad (7)$$

Dimana:

min (x) = Nilai minimum data sampel

max (x) = Nilai maximum data sampel

## II. METODE PENELITIAN

Penelitian adalah sebuah kegiatan yang bertujuan untuk membuat kontribusi orisinal terhadap ilmu pengetahuan (Dawson, 2009). Menurut (Dawson, 2009), terdapat empat metode penelitian yang umum digunakan, diantaranya: action research, experiment, case study dan survey. Penelitian ini menggunakan penelitian eksperimen, yaitu penelitian yang melibatkan penyelidikan perlakuan pada parameter atau variabel tergantung dari penelitiannya dan menggunakan tes yang dikendalikan oleh peneliti itu sendiri.

**Pengumpulan Data (*Data Gathering*)**

Pada bagian ini dijelaskan tentang bagaimana dan darimana data dalam penelitian ini didapatkan. Pada tahap ini ditentukan citra mana saja yang akan diproses.

**Pengolahan Awal Data (*Data Pre-processing*)**

Pengolahan awal data meliputi digitalisasi citra, akuisisi citra dan pembagian data Training dan Testing.

**Metode yang Diusulkan (*Proposed Method*)**

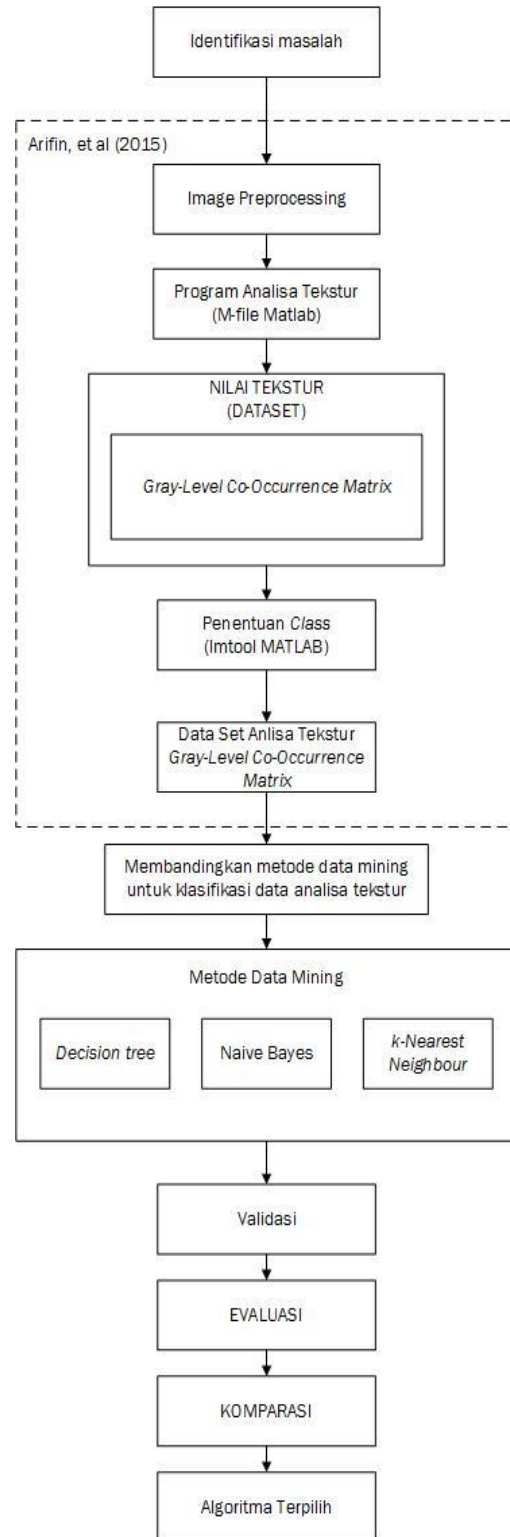
Setelah melakukan pengumpulan dan pengolahan data, maka tahap selanjutnya adalah menentukan metode, dimana metode ini merupakan rangkaian kegiatan untuk melakukan penelitian dan pengujian terhadap data.

**Eksperimen dan Pengujian Metode (*Method Test and Experiment*)**

Metode pengujian model diusulkan pada model yang akan diuji untuk melihat hasil berupa rule yang akan dimanfaatkan dalam mengambil keputusan hasil penelitian. Sebagian data digunakan sebagai data Training dan sebagian lagi sebagai data checking.

**Evaluasi dan Validasi Hasil (*Result Evaluation and Validation*)**

Tahap evaluasi merupakan tahap yang terakhir dari kegiatan penelitian, dimana dalam tahap ini hasil dari tahapan eksperimen akan dievaluasi.



**Gambar 2.1 Metode Penelitian**

**A. PEMBAHASAN**

Penelitian ini bertujuan untuk melakukan perbandingan dan evaluasi pada algoritma klasifikasi data mining.

Penelitian ini menggunakan data analisa tekstur *Gray Level Co-occurrence Matrix* yang terdiri dari 16 atribut diantaranya adalah *Contrast 0, Contrast 45, Contrast 90, Contrast 135, Correlation 0, Correlation 45, Correlation 90, Correlation 135, Energy 0, Energy 45, Energy 90, Energy 135, Homogeneity 0, Homogeneity 45,*

*Homogeneity, 90, Homogeneity 135.* Seluruh atribut akan diseleksi dengan cara memilih atribut berdasarkan derajat keabuan (sudut arah) yang terbaik dan membandingkan dengan menggunakan metode data mining. Di bawah ini adalah data analisa tekstur yang digunakan.

Tabel 4.1 data analisa teksur *Gray Level Co-occurrence Matrix*

Contrast 0	Contrast 45	Contrast 90	Contrast 135	Correlati on 0	Correlati on 45	Correlati on 90	Correlati on 135
0.50328 6	0.55161 5	0.43745 6	0.84867 7	0.97264 8	0.97003 7	0.97624 4	0.95390 1
0.43579 9	0.39503 4	0.57038 9	1.01045 2	0.96705 8	0.96948 1	0.95679 9	0.92284 3
0.38422 5	0.26314 2	0.35826 3	0.77114 2	0.95923 6	0.97139 3	0.96189 1	0.91590 9
1.19027 7	1.48288 0	0.82116 4	1.55646 0	0.94495 2	0.93140 7	0.96204 1	0.92800 5
0.49194 8	0.60117 0	0.56055 6	0.89715 7	0.94241 7	0.92646 9	0.93841 6	0.89955 6
1.80609 418	1.60818 713	0.88888 888	2.04678 362	0.83325 466	0.84605 543	0.92525 018	0.82033 256

Tahapan selanjutnya dalam penelitian ini adalah uji coba dengan menggunakan 3 metode data mining yang berbeda yaitu *Decision tree (C4.5), Naive Bayes, k-Nearest Neighbour.* Dalam eksperimen yang dilakukan data sel analisa tekstur dikelompokkan berdasarkan derajat keabuan dari metode analisa tekstur, diantaranya, all atribut, 0°, 45°, 90° dan 135°.

Tabel 4.2 tabel komparasi hasil pengujian

kelompok atribut	C45	Naive Bayes	k-Nearest Neighbour
all atribut	97,56%	85,31%	95,70%
0°	97,45%	91,73%	96,09%
45°	97,45%	90,89%	96,14%
90°	97,44%	91,26%	95,97%
135°	97,34%	89,98%	96,01%

Di lihat dari hasil komparasi metode data mining untuk klasifikasi sel nukleus dan sel radang maka nilai akurasi yang paling besar terdapat pada metode data mining klasifikasi *Decision tree (C4.5)* dan untuk kelompok atribut yang mempunyai akurasi tertinggi dibandingkan dengan menggunakan

metode data mining lain adalah semua atribut dengan akurasi 97,5148%

*machine, neural network* dan lain-lain.

## B. PENUTUP

### 1. Kesimpulan

Berdasarkan analisa dan pembahasan yang sudah dijelaskan diatas maka dapat diambil kesimpulan sebagai berikut:

- Pada penelitian ini menggunakan data analisa tekstur GLCM (*Gray Level Co-occurrence Matrix*) yang diambil dari penelitian sebelumnya.
- Metode data minning klasifikasi *Decision tree* (C4.5) yang menggunakan semua atribut memiliki akurasi paling tinggi yaitu 97,56%
- Untuk metode data mining klasifikasi Naïve bayes akurasi tertinggi didapat dari kelompok atribut 45° dengan akurasi 90,89%
- Metode data mining *k-Nearest Neighbour*, pada kelompok attribute 90° memiliki akurasi tertinggi sebesar 95,97%.
- Maka setelah dibandingkan metode terbaik untuk klasifikasi analisa tekstur dengan menggunakan GLCM (*Gray Level Co-occurrence Matrix*) terdapat pada metode data mining *Decision tree* (C4.5) dengan kelompok atribut 45°.

### 2. Saran

Diketahui bahwa metode data mining *Decision tree* (C4.5) memiliki akurasi terbaik dibandingkan dengan metode data mining lain yaitu *Naïve bayes* dan *k-Nearest Neighbour* dalam klasifikasi analisa tekstur GLCM (*Gray Level Co-occurrence Matrix*). Namun untuk penelitian selanjutnya hal-hal berikut dapat ditambahkan agar dapat menghasilkan metode yang lebih baik lagi.

- Menerapkan algoritma terpilih dengan aplikasi untuk memudahkan ahli dalam pendeteksian sel nucleus dan sel radang.
- Menerapkan metode data mining lain untuk klasifikasi, yaitu dengan menggunakan *support vectore*

## REFERENCE

- Alfisahrin C4.5, *Naive Bayes* dan *Neural Network*, S. (2014). *Komparasi Algoritma Untuk Memprediksi Penyakit Jantung*. Jakarta: Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- Al-Naggar, R. A. (2012). *Population Health and Preventive Medicine Department, Faculty of Medicine, University Teknologi MARA (UiTM), Malaysia. In Cervical Cancer: Prevention and Control*.
- Arifin, T., Riana, D., & Hapsari, G. I. (2013). Klasifikasi Statistikl Tekstur Sel *Pap Smear* Dengan *Decision Tree*. *Jurnal Informatika*. Bandung: Universitas BSI Bandung.
- Arifin, T., & Riana, D. (2015). Metode Analisa Tekstur Untuk Klasifikasi Sel Nukleus Dan Sel Radang Pada Citra Mikroskopik *Pap Smear*. Jakarta: STMIK Nusa Mandiri Jakarta.
- Ayres, F., Rangayyan, R., & Desautels, J. (2011). *Analysis of Oriented Texture with Applications to the Detection of Architectural Distortion in Mammograms*.
- Bramer, M. (2013). *Principle of Data Mining Second Edition*. London: Springer.
- Bruni, L., Barrionuevo-Rosas, L., Albero, G., Aldea, M., Serrano, B., Valencia, B., et al. (2015). *Human Papillomavirus and Related Diseases Report. In Human Papillomavirus and Related Diseases Report* (pp. 4-7). Barcelona, Spain: *ICO Information Centre on HPV and Cancer (HPV Information Centre)*.
- Burger, W., & Burge, M. J. (2007). *Digital Image Processing*.



- CancerHelp*. (2015). Retrieved July 1, 2015, from <http://www.cancerhelp.org/>
- Dawson, C. (2009). *Projects in Computing and Information Systems*. London: Addison Wesley.
- Fauzi, M., & Tjandrasa, H. (2010). *Implementasi Thresholding Citra Menggunakan Algoritma Hybrid Optimal Estimation*.
- Fatima, M., & Seenivasagam, V. (2012). A hybrid image segmentation of cervical cells by bi-group enhancement and scan line filling. *International Journal of Computer Science and Information Technology & Security (IJCSITS)*, 368-375.
- Hassan, A., & Pei, Z. (2015). *Pap Smear Images Classification for Early Detection of Cervical Cancer*. *International Journal of Computer Applications*, 10-17.
- Herliana, A., & Riana, D., (2013). Klasifikasi Sel Tunggal Pap Smear Berdasarkan Analisis Fitur dan Analisis Tekstur Terseleksi Menggunakan CFS Berbasis *Decision Tree* J48. STMIK Nusa Mandiri. Jakarta.
- Jantzen, J., Norup, G.J., Dounias., & Bjerregaard, B., (2005). *Pap-smear Benchmark Data For Pattern Classification*, Technical University of Denmark, 1-20.
- Kementrian Kesehatan. (2015). Retrieved July 23, 2015, from <http://www.depkes.go.id/>
- livescience*. (2015). Retrieved July 14, 2015, from <http://www.livescience.com/>
- Matrix Laboratory*. (2015). Retrieved July 5, 2015, from <http://www.mathworks.com/>
- Marina, E., & Christophoros, N. (2009). *Automated segmentation of cell nuclei in Pap Smear images*. *IEEE*.
- Marina, E., & Christophoros, N. (2010). *Accurate localization of cell nuclei in Pap Smear images using gradient vector flow deformable models*. *IEEE*.
- Marina, E., & Christophoros, N. (2011). *Accurate Localization Of Cell Nuclei In Pap Smear Images Using Gradient Vector Flow Deformable Models*. Department of Computer Science, University of Ioannina, Ioannina, Greece .
- Marina, E., & Christophoros, N. (2012). *Overlapping Cell Nuclei Segmentation Using a Spatially Adaptive Active Physical Model*. *IEEE Transactions On Image Processing*, 4568-4580.
- Marina, E., & Nikou, C. (2012). *Overlapping Cell Nuclei Segmentation Using a Spatially Adaptive Active Physical Model*. *IEEE*.
- Marina, E., Christophoros, N., & Charchanti, A. (2011). *Automated Detection of Cell Nuclei in Pap Smear Images Using Morphological Reconstruction and Clustering*. *IEEE Transactions On Information Technology In Biomedicine*, 233-241.
- Martin, E. (2003). *Pap-Smear Classification*. Technical University of Denmark. Diambil dari: <http://labs.fme.aegean.gr/decision/downloads/> (25 July 2015).
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. Springer.
- Muhimmah, I., Kurniawan, R., & Indrayanti. (2012). *Automated Cervical Cell Nuclei Segmentation Using Morphological Operation and Watershed Transformation*. *IEEE*, 163-168.
- Muhimmah, I., Kurniawan, R., & Indrayanti. (2013). *Analysis of Features to Distinguish Epithelial Cells and Inflammatory Cells in Pap Smear Images*. *International Conference on Biomedical Engineering and Informatics (BMEI- IEEE)*, 519-523.

- Moshavegh, R., & Ehteshami, B. B. (2013). *Chromatin pattern analysis of cell nuclei for improved cervical cancer screening*. Gothenburg, Sweden.
- Nanni, L., Brahnam, S., Ghidoni, S., Menegatti, E., & Barrier, T. (2013). *Different Approaches for Extracting Information from the Co-Occurrence Matrix*. National Cancer Institute. (2015). Retrieved July 1, 2015, from [www.cancer.gov](http://www.cancer.gov/): <http://www.cancer.gov/>.
- Nurtanio, I., Astuti, E. R., Purnama, E. I., Hariadi, M., & Purnomo, M. H. (2013). *Classifying Cyst and Tumor Lesion Using Support Vector Machine Based on Dental Panoramic Images Texture Features*. IAENG International Journal of Computer Science.
- Oscanoa, J., Mena, M., & Kemper, G. (2015). *A Detection Method of Ectocervical Cell Nuclei for Pap test Images, Based on Adaptive Thresholds and Local Derivatives*. International Journal of Multimedia and Ubiquitous Engineering, 37-50.
- Prasetyo, E. (2011). *Pengolahan citra digital dan aplikasinya menggunakan Matlab*. Yogyakarta: Penerbit Andi.
- Pratama, G. K., Riana, D., & Hasanudin. (2012). *Pap Smear Nuclei Texture Analysis*. International Conference on Women's Health in Science & Engineering. Bandung: Institut Teknologi Bandung, 1-4.
- Rangayyan, R., Nguyen, T., Ayres, F., & Nandi, A. (2010). *Effect of Pixel Resolution on Texture Features of Breast Masses in Mammograms*.
- Riana, D., Marina, E., Christophoros, N., Dwi, H., Tati Latifah, R., & Kalsoem, O. (2015). *Inflammatory Cell Extraction and Nuclei Detection in Pap Smear Images*. International Journal of E-Health and Medical Communications, 27-43.
- Riana, D. (2010). *Hierarchical Decision Approach Berdasarkan Importance Performance Analysis Untuk Klasifikasi Citra Tunggal Pap Smear Menggunakan Fitur Kuantitatif dan Kualitatif*.
- Smith, J. S., Lindsay, L., Hoots, B., Keys, J., Franceschi, S., Winer, R., et al. (2007). *Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: A meta-analysis update*. the International Union Against Cancer.
- Sreedevi, M., Usha, B., & Sandya, S. (2012). *Papsmear Image based Detection of Cervical Cancer*. International Journal of Computer Applications.
- Sokouti, B., & Haghypour, S. (2011). *A Pilot Study on Image Analysis Techniques for Extracting Early Uterine Cervix Cancer Cell Features*. Springer Science+Business Media, LLC.
- Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*.
- Tareef, A., Yang, S., Weidong, C., David, D., Feng, & Mei, C. (2014). *Automated Three-Stage Nucleus and Cytoplasm Segmentation of Overlapping Cells*. IEEE.
- Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. Cornwall: John Wiley & Sons, Ltd.
- Vistekdatabase. (2015). Retrieved July 23, 2015, from <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>
- WHO. (2015). Retrieved June 1, 2015, from [WHO: http://www.who.int/en/](http://www.who.int/en/)
- Witten, I., & Frank, E. (2005). *Data Mining: Practical Machine*



*Learning Tools and Techniques  
2nd Edition. USA: Elsevier.*

Xu, D.-H., Kurani, A., Furst, J., &  
Raicu, D. (2004). *Run-length  
encoding for volumetric texture.  
School of Computer Science,  
Telecommunications, and  
Information Systems, DePaul  
University, Chicago, Illinois.  
USA.*