

KLASIFIKASI SEL TUNGGAL P_{AP} SMEAR BERDASARKAN ANALISIS FITUR BERBASIS NAÏVE BAYES CLASSIFIER DAN PARTICLE SWARM OPTIMIZATION

Taufik Hidayatulloh¹, Asti Herliana², Toni Arifin³

¹ Program Studi Manajemen Informatika
AMIK BSI Jakarta
Jl. RS. Fatmawati No. 24 Pondok Labu, Jakarta Selatan
e-mail: taufik.tho@bsi.ac.id

² Program Studi Sistem Informasi
Universitas BSI Bandung
Jl. Sekolah Internasional No.1-6 Antapani, Bandung
e-mail: asti.ala@bsi.ac.id

³ Program Studi Teknik Informatika
Universitas BSI Bandung
Jl. Sekolah Internasional No.1-6 Antapani, Bandung
e-mail: toni.tfn@bsi.ac.id

ABSTRACT

Research from the informatics experts about cervical cancer mainly single cell of the Pap smear, increasingly showing the almost perfect results. 20 features produced by research conducted by Jantzen, Norup, Dounias and Bjerregaard, has now been developed and reviewed. This assessment takes precedence on efficiency features that make a significant contribution (assessed based on the percentage of best feature tool). Until now, the problems that have not been able to solve is to maximize the results of the classification of the 7th grade single cells of Pap Smear. This is due to the lack of research experts with a combination of the best methods that produce maximum results. After reviewing previous studies, classification methods that provide the best value to date is Naive Bayes. For the optimization method used in the present study is the Particle Swarm Optimization. With a combination of methods Naive Bayes and Particle Swarm Optimization, obtained better results from previous research that is 62.67% for the classification of 7 classes and 95.70% for the classification of 2 classes.

Keyword : *Single Cell, Pap Smear, Feature Analysis, Naive Bayes, Particle Swarm Optimization*

I. PENDAHULUAN

Salah satu penyakit penyebab kematian pada perempuan adalah kanker serviks. Kanker ini menyerang bagian leher rahim dan disebabkan oleh *Human Papilloma Virus* (HPV). Gejala dari penyakit ini tidak langsung terasa, akan tetapi akan mulai menunjukkan tanda-tandanya sekitar 10-20 tahun setelah pertama kali terserang. Begitu ganasnya penyakit ini, menyebabkan banyaknya studi yang dilakukan oleh para ahli patologi agar dapat menanggulangi penyakit ini. Untuk menanggulangi dan mendeteksi keberadaan virus HPV ini sejak awal, para ahli patologi menemukan metode *Pap Smear* sebagai sebuah metode deteksi awal keberadaan virus HPV. Dengan metode *Pap Smear* keberadaan virus

HPV dapat terlihat dengan adanya perubahan bentuk nukleus.

Sel tunggal *Pap Smear* pertama kali diperkenalkan oleh Erik Martin pada tahun 2003. Penelitian yang dilakukan oleh Erik Martin pada tahun 2003 didasarkan pada hasil sel tunggal yang didapatkan melalui metode *Pap Smear* yang ditemukan oleh Georgius Papanicolau pada tahun 1930 dengan ditemukannya mekanisme diagnosa pra kanker rahim yang pada awal kemunculannya dikenal dengan nama metode Papanicolau (Riana, 2010). Dari penelitian yang dilakukan oleh Erik Martin ini, didapatkan hasil berupa klasifikasi 7 kelas dari sel tunggal *Pap Smear* dimana didalamnya didapatkan klasifikasi yakni 2 kelas yaitu kelas normal dan kelas abnormal (Martin, 2003).

Pada penelitian selanjutnya yang dilakukan oleh J.Jantzen, J.Norup, G.Dounias dan B.Bjerregaard pada tahun 2005, dilakukan segmentasi dan klasifikasi dari 917 citra sel *Pap Smear* dengan 20 fitur yang diklasifikasikan kedalam 7 kelas (3 kelas normal dan 4 kelas abnormal) yang sekarang dikenal dengan data herlev (Jantzen, Norup, Dounias, & Bjerregaard, 2005), yang dapat digunakan untuk menguji beberapa metode klasifikasi serta kombinasi diantaranya untuk mendapatkan metode terbaik untuk proses klasifikasi sel *Pap Smear*. Perkembangan penelitian dari data herlev ini, sudah menunjukkan hasil yang cukup memuaskan dengan menggunakan metode seperti *Decision Tree*(J48) dan *Hierarchical Decision Approach* (HDA) serta analisa *Importance Performance Analysis*.

Meskipun hasil klasifikasi terhadap 7 kelas dan 2 kelas dari sel tunggal *Pap Smear* sudah cukup memuaskan, namun tingkat akurasi masih belum 100%. Diantara metode yang menunjukkan hasil terbaik dengan set data fitur herlev ini adalah metode *Hierarchical Decision Approach* (HDA) dan analisa *Importance Performance Analysis* dengan hasil akurasi yang didapatkan yakni 83,21% (Riana, 2010), *Decision Tree* dan Optimasi menggunakan *Correlation based Features Selection* (CFS) menghasilkan sensitifitas mencapai 97.05% dan 98.90% untuk performa sistemnya (Muhimmah, Anwariyah & Indrayanti, 2012), Metode J48 dengan hasil performa didapatkan klasifikasi terbaik sebesar 73,83%. Dengan melihat besarnya data yang dimiliki oleh data set sel tunggal *Pap Smear*, maka dibutuhkan metode *Naïve Bayes Classifier* sebagai metode klasifikasinya dan optimasi menggunakan *Particle Swarm Optimization* (PSO) untuk meningkatkan akurasi hasil klasifikasinya.

Melihat banyaknya penelitian yang membahas mengenai sel tunggal ini, maka dapat dikatakan penelitian ini penting untuk dilakukan. Hal ini untuk terus meningkatkan hasil penelitian sehingga hasil klasifikasi dari sebuah sel tunggal mencapai akurasi sebesar 100%. Jika hasil yang didapatkan mendekati 100%, maka akan membantu ahli patologi untuk menentukan apakah pasien tersebut terkena kanker serviks atau tidak serta dapat membantu para ahli patologi untuk memberikan perawatan yang tepat.

II. KAJIAN LITERATUR

Tinjauan penelitian terdahulu untuk objek sel tunggal *Pap Smear* dijabarkan sebagai berikut:

1. Pada tahun 2003, Martin melakukan klasifikasi terhadap sel *Pap Smear* dan membaginya kedalam 2 kelas yakni kelas normal dan abnormal dan 7 kelas yakni *Superficial epithelial*, *Intermediate epithelial*, *Columnar epithelial*, *Mild dysplastic*, *Moderate dysplastic*, *Severe dysplastic*, dan *Carcinoma in situ* dengan berdasarkan kepada 20 fitur. Pada penelitian ini didapatkan tingkat *correctness* untuk klasifikasi 2 kelas rata-rata sudah diatas 90%.
2. Jantzen, Norup, Dounias dan Bjerregaard pada tahun 2005 melakukan penelitian terhadap 917 sampel citra sel tunggal pap smear dan menghasilkan basis data dari 917 citra sel tunggal *Pap Smear* kedalam klasifikasi menjadi 2 kelas yakni normal dan abnormal serta menjadi 7 kelas yakni *Normal Superficial*, *Normal Intermediate*, *Normal Columnar*, *Mild (Light) Dysplasia*, *Moderate Dysplasia*, *Severe Dysplasia*, dan *Carcinoma In Situ*. Basis data yang dihasilkan dari penelitian ini kemudian kita kenal dengan data set citra sel tunggal Herlev.
3. Pada tahun 2008 Mustafa, Isa, Mashor dan Othman melakukan penelitian terhadap sel kanker serviks dengan pendekatan fitur ekstraksi seperti *perimeter*, *red*, *green*, *blue*, *intensity1*, *intensity2* dan *saturation*. Dari penelitian ini didapatkan tingkat akurasi mencapai 97.65 % untuk data train, 90.91% untuk data test dan 94.29% untuk data keseluruhan.
4. Pada tahun 2010 Riana, Murni, Widyantoro dan Mengko melakukan penelitian terhadap sel tunggal citra *Pap Smear* menggunakan fitur kuantitatif dan kualitatif sebagai fitur yang diteliti dan metode *Hierarchical Decision Approach* (HDA) dan analisa *Importance Performance Analysis* sebagai metode klasifikasinya. Dari penelitian ini didapatkan hasil rata-rata akurasi sebesar 83,206% terhadap lima kelas. Masih juga terdapat kesulitan pengklasifikasian terhadap kelas *Moderate Dysplasia* dan *Severe Dysplasia*.
5. Pada tahun 2012 Pratama, Riana, Dewi, Widyantoro, dan Mengko melakukan penelitian terhadap sel citra tunggal *Pap Smear*. Pada penelitian kali ini didapatkan 18 parameter tekstur.
6. Masih ditahun 2012, Muhimmah, Anwariyah dan Indrayanti melakukan penelitian dari sel kanker serviks dengan ekstraksi dan fitur seleksi yang digunakan

adalah *shape area, shape factor, roundness, eccentricity, normalized, third moment, uniformity, entropy* dan *mean histogram*. Dari hasil penelitian ini didapatkan tingkat sensitifitasnya mencapai 97.05% dan 98.90% untuk performa sistemnya.

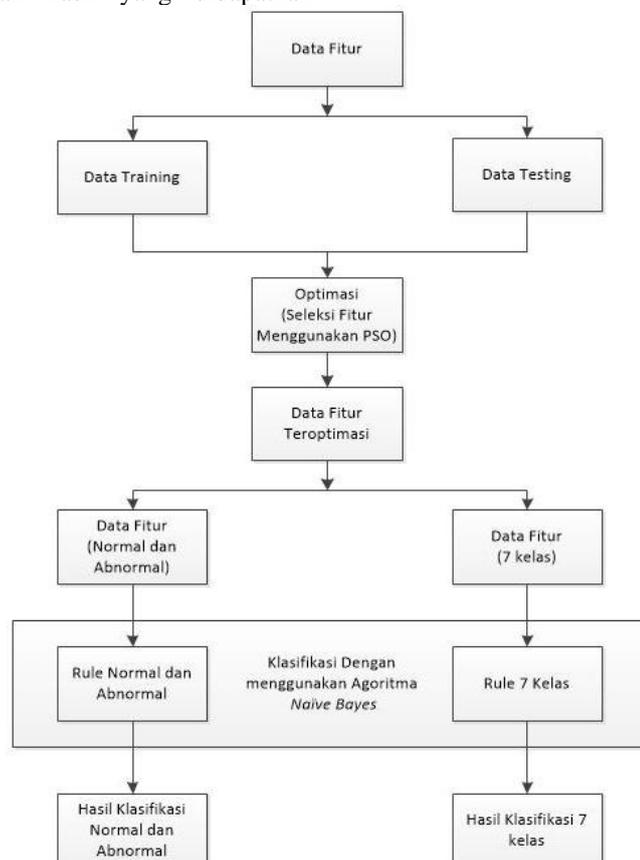
7. Pada tahun 2013 Arifin, Riana dan Hapsari melakukan penelitian lanjutan dari Pratama, Riana, Dewi, Widyantoro, dan Mengko. Pada penelitian ini ditambahkan 1 fitur *brightness*, sehingga 18 parameter tekstur sebelumnya yang telah ditemukan bertambah menjadi 19 parameter. Selain menambahkan parameter, pada penelitian ini juga didapatkan prosentase *correctness* dengan menggunakan *Gray Level Co-occurrence Matrices (GLCM)* untuk analisis teksturnya dimana didapatkan prosentase 73,6% untuk klasifikasi 2 kelas (normal dan abnormal) dan 34,6% untuk klasifikasi 7 kelas (*Normal Superficial, Normal Intermediate, Normal Columnar, Mild (Light)Dysplasia, Moderate Dysplasia, Severe Dysplasia, Carcinoma In Situ*)
8. Pada tahun 2013 Riana, Widyantoro dan Mengko melakukan penelitian tentang tekstur sel nuklesu dari sel tunggal *Pap Smear* dengan hasil yang didapatkan

adalah bahwa dengan menggunakan algoritma J48 didapatkan klasifikasi performa terbaik sebesar 73,8277%. Selain itu, dari penelitian ini didapatkan juga kesimpulan bahwa dari 5 parameter yang digunakan dalam penelitian ini yakni *entropy, contrast, correlation, energy* dan *homogeneity* didapatkan grafik yang bertumpuk untuk parameter *energy* yang berarti hasil ekstraksi dari parameter ini adalah konsisten dan tidak ada perbedaan nilai untuk arah $0^0, 45^0, 90^0$ dan 135^0 .

9. Pada tahun 2013 Herliana dan Riana melakukan penelitian mengenai optimasi klasifikasi sel tunggal *Pap Smear* menggunakan metode klasifikasi *Decision Tree J48* dan *Clasification based Features Selection (CFS)*. Hasil dari penelitian ini berupa peningkatan akurasi sebesar 90% untuk 2 kelas dan akurasi sebesar 67,87% untuk 7 kelas.

III. METODE PENELITIAN

Dari jenis penelitian yang telah ditentukan sebelumnya, diperlukan model penelitian untuk menggambarkan *Rule* penelitian dari awal hingga mendapaktan hasil akhir. Berikut adalah model penelitian yang dilakukan.



Gambar 1. Metodologi Penelitian

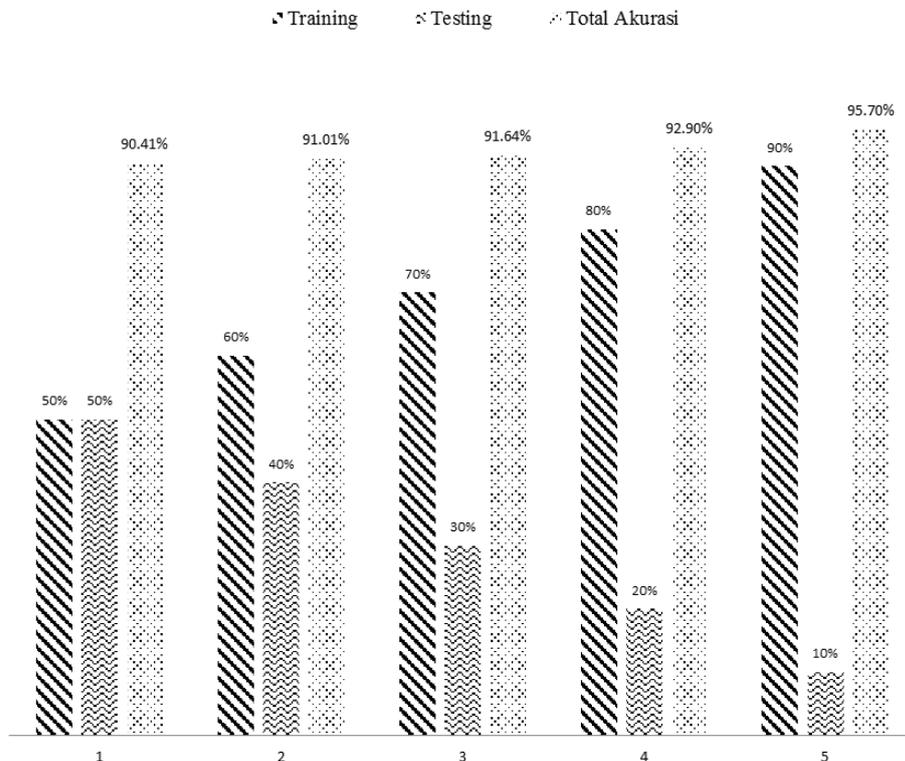
IV. PEMBAHASAN

Dari hasil pengujian model klasifikasi Naïve Bayes lalu di optimasi dengan menggunakan metode *Particle swarm optimization*, diketahui bahwa hasil akurasi yang didapatkan beragam, hasil tersebut didapatkan dengan melakukan ujicoba data *training* dan *testing* dengan jumlah data yang berbeda. Pada percobaan yang telah dilakukan nilai akurasi untuk klasifikasi normal dan abnormal yang paling tinggi didapatkan dari percobaan dengan data *training* berjumlah 826

data (90% dari 917 total data) dan data *testing* sebanyak 92 data (10% dari 917 total data) dengan akurasi 95,70%, sedangkan untuk akurasi klasifikasi kelas 1 sampai dengan kelas 7 nilai akurasi paling tinggi didapatkan dari percobaan dengan data *training* 734 data (80% dari 917 total data) dan data *testing* sebanyak 183 data (20% dari 917 total data) dengan akurasi 62,67%. Di bawah ini adalah tabel dan grafik hasil pengujian model.

Tabel 1. Hasil Pengujian untuk Kelas Normal dan Abnormal

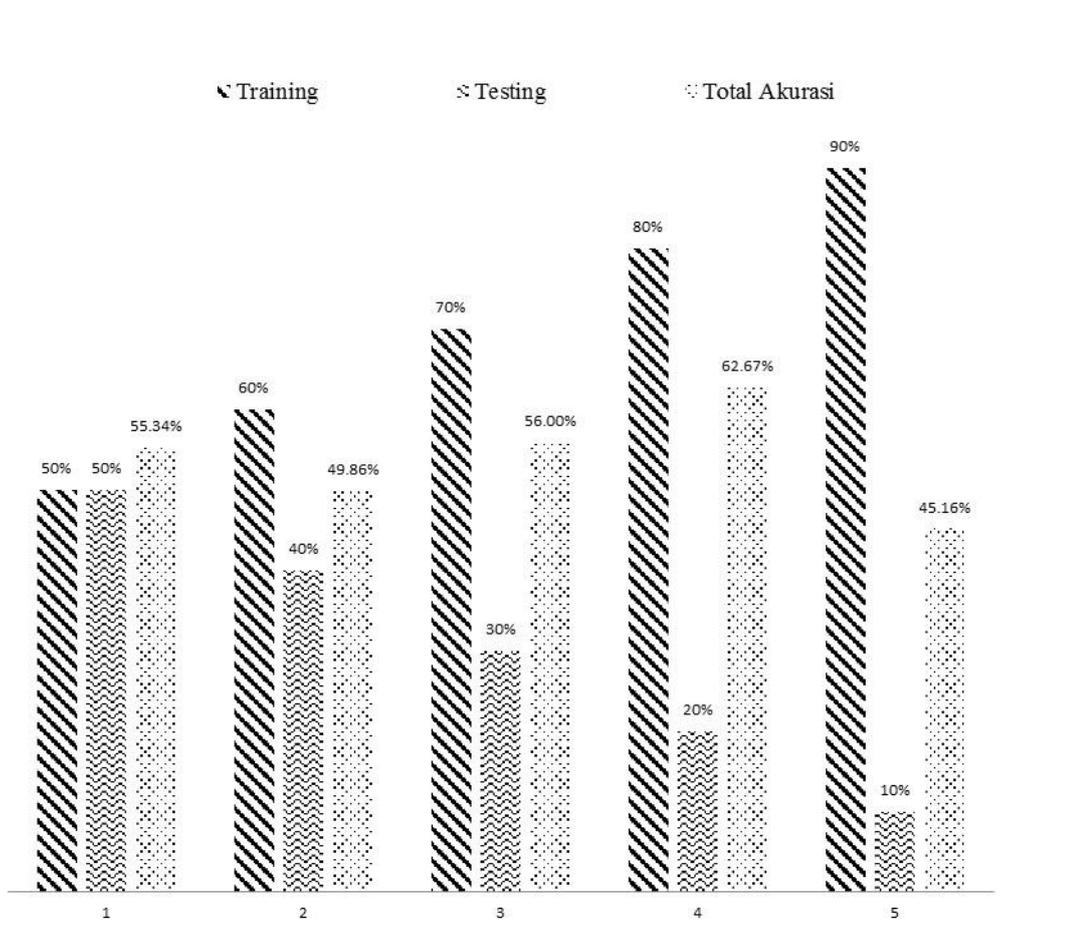
Algoritma	Training	Testing	Total Akurasi
NBC-PSO	50%	50%	90.41%
NBC-PSO	60%	40%	91.01%
NBC-PSO	70%	30%	91.64%
NBC-PSO	80%	20%	92.90%
NBC-PSO	90%	10%	95.70%



Gambar 2. Grafik hasil pengujian untuk kelas Normal dan Abnormal

Tabel 2. Hasil Pengujian untuk Kelas 1 sampai dengan kelas 7

Kelas 1 -7			
Algoritma	Training	Testing	Total Akurasi
BC-PSO	50%	50%	55.34%
NBC-PSO	60%	40%	49.86%
NBC-PSO	70%	30%	56.00%
NBC-PSO	80%	20%	62.67%
NBC-PSO	90%	10%	45.16%



Gambar 3. Grafik hasil pengujian untuk Kelas 1 sampai dengan kelas 7

V. PENUTUP

Kesimpulan

Berdasarkan hasil analisa dan pembahasan yang sudah di paparkan di atas maka dapat diambil kesimpulan sebagai berikut:

- a. ada penelitian ini menggunakan data yang didapat dari *Pap smear Benchmark Data For Pattern*

Clasification J. Jantezen l, J. Norup, G. Dounias, and B. Bjerregaard, dari University Hospital Dept. of Pathology Herlev (917 data).

- b. Hasil dari analisa dan percobaan yang sudah dilakukan untuk klasifikasi kelas normal dan abnormal yang paling tinggi

didapatkan dari data *training* 826 data (90%) dan *testing* 92 (10%).

- c. Klasifikasi kelas 1 sampai dengan kelas 7 akurasi yang paling tinggi didapatkan dari data *training* 734 (80%) dan *testing* 183 (20%).

Saran

Diketahui bahwa hasil akurasi sel normal dan abnormal lebih besar dibandingkan dengan hasil akurasi untuk klasifikasi kelas 1 sampai dengan 7 maka dari itu untuk penelitian selanjutnya hal-hal berikut dapat ditambahkan agar dapat menghasilkan akurasi yang lebih baik lagi, diantaranya sebagai berikut;

- Menerapkan metode lain untuk klasifikasi, yaitu dengan menggunakan metode *Support vectore machine, neural network, Genethic algorithm* dan lain-lain.
- Menerapkan algoritma terpilih dengan membuat aplikasi untuk memudahkan ahli patologi dalam pengamatan sel *Pap Smear*.

REFERENSI

- Alfisahrin, S. (2014). Komparasi Algoritma C4.5, Naive Bayes dan Neural Network 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 Statistik Tekstur Sel Pap Smear Dengan Decision Tree. *Jurnal Informatika*. Bandung: Universitas BSI Bandung.
- Arifin, T. (2014). Implementasi Metode K-Nearest Neighbor Untuk Klasifikasi Citra Sel Pap Smear Menggunakan Analisis Tekstur Nukleus. *Jurnal Informatika, Universitas BSI Bandung*.
- Arifin, T. (2014). Metode Data Mining Untuk Klasifikasi Data Sel Nukleus Dan Sel Radang Berdasarkan Analisa Tekstur. *Jurnal Informatika, Universitas BSI Bandung*.
- 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.
- 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., & Nikou, C. (2012). Overlapping Cell Nuclei Segmentation Using a Spatially Adaptive Active Physical Model. *IEEE*.
- 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.
- 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/>.
- 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.
- 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 juli 23, 2015, from <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vis tex.html>
- WHO. (2015). Retrieved juny 1, 2015, from WHO: <http://www.who.int/en/>
- Witten, I., & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques 2nd Edition. USA: Elsevier.