

Klasifikasi Statistik Tekstur Sel Pap Smear Dengan Decesion Tree

Toni Arifin, Dwiza Riana, Gita Indah Hapsari

Teknik Informatika

Universitas BSI

Jl. Sekolah Internasional No. 1-6, Bandung 40282, Indonesia

toni.tfn@bsi.ac.id, dwiza@bsi.ac.id

ABSTRAK

Penelitian ini menyajikan analisis tekstur dan klasifikasi citra sel pap smear. Pada analisis tekstur difokuskan pada citra nukleus sel Pap smear, metode yang digunakan adalah metode Gray Level Co-occurrence Matrix (GLCM) dengan menggunakan lima parameter yaitu korelasi, energi, homogenitas dan entropi ditambah dengan menghitung nilai Brightness pada citra yang diproses. Citra yang digunakan dalam penelitian ini menggunakan data citra Harlev, yang terdiri dari 280 citra yang sudah dikategorikan ke dalam 7 kelas yaitu 3 kelas sel normal yang meliputi Normal Superficial, Normal Intermediate, and Normal Columnar dan 4 kelas lainnya adalah kategori kelas citra sel abnormal yang meliputi Mild (Light) Dyplasia, Moderate Dysplasia, Severe Dysplasia dan Carcinoma In Situ. Berdasarkan hasil pengolahan citra yang menghasilkan nilai matriks dari setiap parameter yang dihitung, citra sel Pap smear akan diklasifikasikan menurut jenisnya normal atau abnormal dan berdasarkan kelasnya dengan menggunakan decision tree yang diolah dengan algoritma clasifier J48 pada aplikasi weka. Untuk akurasi yang dihasilkan dari klasifikasi sel normal dan abnormal adalah 73% dan untuk akurasi klasifikasi tujuh kelas adalah 34,3%.

Kata Kunci : Klasifikasi, Statistik Tekstur, Sel Pap Smear, *Decision Tree*.

ABSTRACT

This research presents the texture analysis and classification of cells pap smear image. Texture analysis focused on the cell nucleus Pap smear image, the research method used the Gray Level Co-occurrence Matrix (GLCM) method, by using five parameter that include contrast, correlation, energy, homogeneity, entropy and brightness. The image used in this research using image data Harlev. The images from 280 subjects are categorized into seven classes. Three classes of which are normal cell image class categories that include Normal Superficial, Normal Intermediate, and Normal Columnar, and the other four classes are categories of abnormal cell image class that include Mild (Light) Dyplasia, Moderate Dysplasia, Severe Dysplasia and Carcinoma In Situ. Based on the results of image processing that produces a matrix of values of each parameter were calculated, Pap smear cell image will be classified according to the type of normal or abnormal and based on the class using the decision tree treated with algorithm clasifier J48 in weka applications. To the resulting accuracy of the classification normal and abnormal cells is 73% and for seven class classification accuracy is 34,3%.

Keywords : Classification, Statistical Texture, Cell Pap Smear, *Decision Tree*

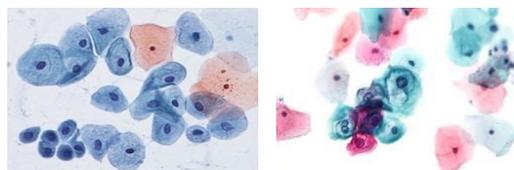
1. PENDAHULUAN

Kanker merupakan penyakit yang berbahaya dan penyebab kematian tertinggi di Indonesia bahkan dunia. Menurut data dari WHO tahun 2012, Penyakit kanker leher rahim atau kanker serviks merupakan kanker yang diderita kaum wanita, kanker ini disebabkan oleh *papilloma virus* http://www.who.int/2013/01/19/reproductive_health/publications/cancers/9789241505147/en/index.html, dan menurut Dalimartha, S. (2004) Pendeteksian kanker serviks secara dini dapat dilakukan melalui beberapa cara salah satunya adalah Pap smear yaitu pemeriksaan usapan mulut rahim untuk melihat sel-sel mulut rahim di bawah mikroskop. Bila dikaitkan dengan perkembangan ilmu pengetahuan dan teknologi, pendeteksian dapat dilakukan melalui pengolahan citra. Pengolahan citra yang dilakukan untuk mendeteksi kanker serviks dapat dilihat dari segi warna dan tekstur sel Pap smear, bila dilihat dari segi warna dapat dilihat dari warna merah dan putih sedangkan bila dilihat dari tekstur dapat dilihat melalui derajat keabuan di dalam sekumpulan pixel-pixel yang saling berdekatan. Analisis tekstur sel Pap smear dapat dilakukan melalui beberapa metode statistik diantaranya adalah GLCM (*Gray level Co-occurrence Matrix*), GLRL (*Gray level run length*), dan LBP (*Liquid Base Preparation*). Tujuan dari penelitian ini untuk mengetahui nilai matrik dari setiap tekstur sel Pap smear agar dapat diketahui jenis sel yang diproses apakah termasuk sel normal atau abnormal dengan melakukan proses klasifikasi. Paper ini terbagi dalam beberapa bagian. Bagian 2 membahas tentang kanker serviks, bagian 3 membahas metode penelitian yang digunakan, bagian 4 membahas tentang material yang digunakan dalam penelitian, dan bagian 5 menjelaskan tentang pembahasan dan hasil penelitian, selanjutnya ditutup dengan kesimpulan dan rencana penelitian lanjutan.

II. KAJIAN LITERATUR

Menurut data dari WHO tahun 2012, Leher rahim merupakan bagian bawah rahim, tempat dimana bayi tumbuh saat kehamilan. Kanker serviks disebabkan oleh jenis virus yang disebut *human papilloma virus* (HPV). Virus dapat menyebar melalui kontak seksual. Selain itu merokok, memiliki banyak anak, menggunakan pil KB dalam jangka

waktu yang lama dan memiliki infeksi HIV merupakan salah satu factor penyebab timbulnya kanker serviks http://www.who.int/2013/01/19/reproductive_health/publications/cancers/9789241505147/en/index.html. Menurut Dalimartha, S. (2004) Pada awal gejala kanker serviks kebanyakan wanita tidak merasakan gejala apapun, kemudian akan mengalami pendarahan pada vagina. Biasanya memerlukan waktu beberapa tahun untuk sel-sel normal pada serviks berubah menjadi sel kanker. Pemeriksaan sel normal dan abnormal dapat dilakukan dengan tes Pap smear. Pap smear adalah pemeriksaan sel leher rahim dibawah mikroskop. Gambar 1 dibawah ini menunjukkan sel normal dan sel yang abnormal.



Gambar 1 Sel Normal dan Sel Abnormal

III. METODE PENELITIAN

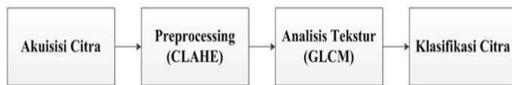
Penelitian ini menggunakan 280 citra. Data dari sel tunggal kanker serviks tersebut telah diklasifikasikan ke dalam 7 kelas, masing kelas mempunyai 40 citra sel. Dari ke 7 kelas tersebut, 3 kelas pertama adalah kelas normal yang terdiri dari *Normal Superficial* (NS), *Normal Intermediate* (NI), dan *Normal Columnar* (NC). Sedangkan 4 kelas berikutnya masuk ke kategori abnormal yaitu *Mild Light Dysplasia* (MLD), *Severe Dysplasia* (SD), *Moderate Dysplasia* (MD), dan *Carcinoma In Situ* (CIS).

Sample yang digunakan dalam penelitian ini terdiri dari 120 citra sel dari kelas normal. Untuk kelas abnormal digunakan 160 citra, tiap kelas sebanyak 40 citra.

Tahapan pertama dalam pendekatan metode penelitian yang diusulkan untuk klasifikasi sel normal dan abnormal Pap smear adalah mengkonversi citra RGB ke *grayscale* tanpa merubah kontras pada citra asli pada tahap ini citra hasil konversi memiliki kontras rendah, sehingga akan mempersulit dalam proses selanjutnya. Pada tahap kedua meningkatkan intensitas kontras hasil konversi *grayscale* dan mempertajam tepi gambar hasil konversi. Pada tahapan ketiga adalah analisis tekstur menggunakan

metode *Gray level Co-occurrence Matrix* (GLCM) metode ini menghasilkan nilai matrik dari setiap tekstur dengan menggunakan 6 parameter diantaranya *Entropy, Contrast, Correlation, Energy, Homogeneity, Deviation* dan *Brightness*.

Tahap terakhir adalah klasifikasi, pada tahap ini nilai matrik dari setiap parameter akan diproses untuk proses klasifikasi sel normal dan abnormal.



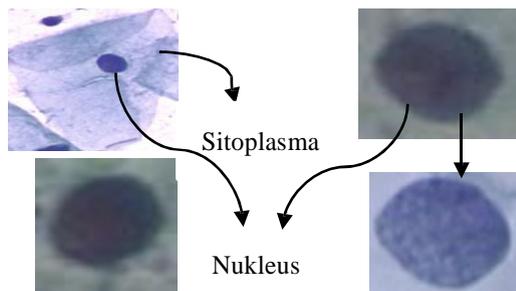
Gambar 2. Blok Diagram

A. Proses Cropping dan Akuisisi Citra

Sample yang digunakan dalam penelitian ini terdiri dari 120 citra sel dari kelas normal. Untuk kelas abnormal digunakan 160 citra, tiap kelas sebanyak 40 citra.

Setiap citra memiliki *size* yang berbeda-beda dan mempunyai warna RGB masing. Pada tahap ini semua citra RGB dikonversi ke *grayscale* menggunakan fungsi `rgb2gray` dengan persamaan sebagai berikut.

$$\text{Gray} = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1)$$

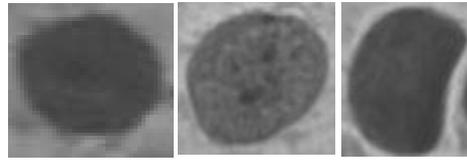


Gambar 3. Ilustrasi Proses Cropping Pada Sel Normal dan Abnormal

B. Preprosesing

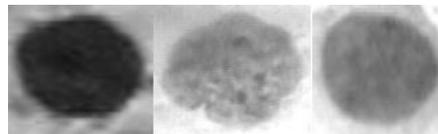
Grayscale pada citra sel Pap smear memiliki nilai kontras yang rendah, sehingga menyebabkan kesulitan dalam analisis tekstur. Preprosesing ini bertujuan untuk meningkatkan kontras dan mempertajam tepi gambar. Pada proses ini digunakan 2 metode yaitu merubah intensitas warna dari citra *Grayscale* (*Image adjustment*) dan

menggunakan metode *Contrast Limited Adaptive Histogram* (CLAHE).



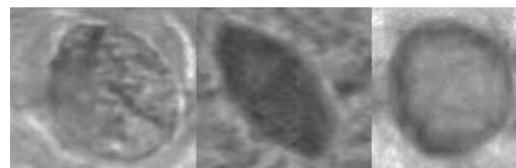
Gambar 4 Citra Hasil Proses *Grayscale*

Image adjustment (`imadjust`) merupakan *basic tool* dari *image processing toolbox* yang disediakan oleh Matlab untuk transformasi intensitas *grayscale* dari sebuah citra, persamaannya adalah sebagai berikut :



Gambar 5 Citra Hasil Proses *adjustment*

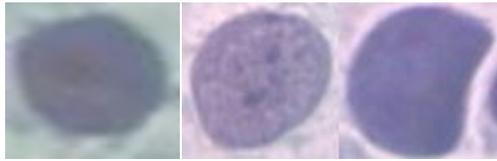
Contrast Limited Adaptive Histogram (CLAHE). Termasuk teknik perbaikan citra yang digunakan untuk memperbaiki kontras pada citra. CLAHE beroperasi pada region yang kecil dan disebut dengan *tile*. Keuntungan menggunakan CLAHE adalah perhitungan yang sederhana, mudah digunakan dan menghasilkan *output* yang baik, karena memiliki sedikit noise yang sedikit.



Gambar 6. Citra Hasil Proses *Contrast Limited Adaptive Histogram* (CLAHE).

C. *Brightness*

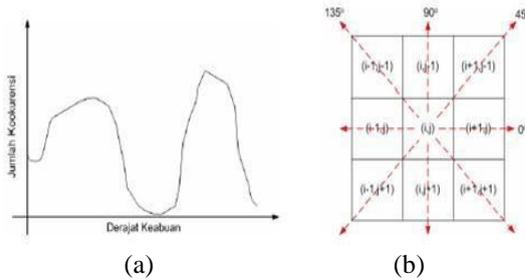
Brightness adalah nama lain dari tingkat kecerahan/intensitas cahaya. Elemen ini menyatakan banyaknya cahaya yang diterima oleh mata. Dalam penelitian ini pengolahan citra *Brightness* bertujuan untuk mengetahui tingkat kecerahan pada citra yang diolah, setelah itu akan muncul nilai matrik, nilai matriks ini akan diolah bersama nilai matriks 5 parameter dan kesemua nilai akan digunakan untuk klasifikasi citra.



Gambar 7 Citra Hasil Proses *Brightness*

D. Analisis Tekstur

Analisis tekstur difokuskan pada citra inti sel Pap smear, dengan menggunakan Matlab. Pada analisis tekstur ini menggunakan metode GLCM (*Gray Level Co-occurrence Matrix*) Ada 5 parameter yang digunakan yaitu: kontras, korelasi, energi, homogenitas dan entropi. Dalam *Co-occurrence matrix* Jarak dinyatakan dalam piksel dan orientasi dinyatakan dalam derajat. Orientasi dibentuk dalam empat arah sudut dengan interval sudut 45°, yaitu 0°, 45°, 90°, dan 135°.



Gambar 8 Contoh Ilustrasi Ekstraksi Ciri Statistik, (a) Histogram citra sebagai fungsi probabilitas kemunculan nilai intensitas pada citra, (b) Hubungan ketetangaan antar piksel sebagai fungsi orientasi dan jarak spasial

Dibawah ini adalah parameter GLCM (*Gray Level Co-occurrence Matrix*) yang digunakan pada penelitian ini.

1) Energi (Energy).
$$\sum \sum \quad (2)$$

2) Entropi (Entropy).
$$\sum \sum \quad (3)$$

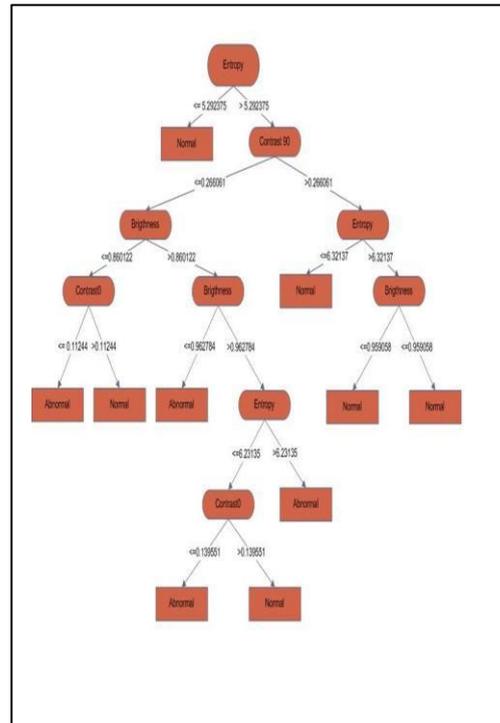
3) Kontras (Contrast).
$$\sum \sum \quad (4)$$

4) Homogenitas (*Homogeneity*).
$$\sum \sum \quad (5)$$

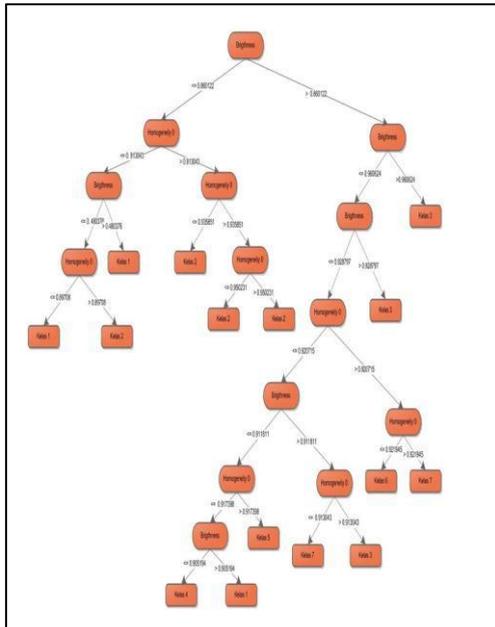
5) Korelasi (Correlation).
$$\sum \sum \quad (6)$$

E. Klasifikasi

Nilai Matriks hasil dari analisa tekstur akan dijadikan acuan untuk mendapatkan rule Klasifikasi normal dan abnormal dan rule klasifikasi kelas. Aplikasi yang digunakan untuk mendapatkan rule klasifikasi adalah aplikasi weka (*Waikato Environment for Knowledge Analysis*) dengan tools *classifiers.j48* untuk mendapatkan rule terbaik digunakan 2 metode yang sudah pada tools *classifiers.j48* yaitu *buildClassifier* dan *classify Instance*.



Gambar 9. *Decision Tree* Klasifikasi Normal dan Abnormal



Gambar 10. *Decision Tree* Klasifikasi Kelas

IV. PEMBAHASAN

Hasil dari proses klasifikasi citra akan diuji dengan pengujian *Confusion Matrix*. Pengujian ini bertujuan untuk menguji seberapa baik sistem klasifikasi yang dibuat.

Tabel 1 Pengujian Klasifikasi Normal dan Abnormal

		Kelas Hasil Prediksi	
		Normal	Abnormal
Kelas Aktual	Normal	54	31
	Abnormal	6	49

Dari hasil perhitungan akurasi klasifikasi normal dan abnormal diatas sistem yang dibuat mempunyai akurasi ketepatan dalam memprediksi sebesar 73,6 % dan kesalahan dalam memprediksi sebesar 26,4 %.

Tabel 2 Pengujian Klasifikasi Kelas

Kelas Aktua 1	Kelas Hasil Klasifikasi						
	Kls 1	Kls 2	Kls 3	Kls 4	Kls 5	Kls 6	Kls 7
Kls 1	8	4		4		2	5
Kls 2	1	11	5	5	3	4	8
Kls 3	7	2	11	1	2	1	
Kls 4		1	2	8	4	3	4
Kls 5	4	1	1	2	5	4	
Kls 6							
Kls 7	2	1	1	3	5	5	5

Dari hasil perhitungan akurasi klasifikasi kelas diatas sistem yang dibuat mempunyai akurasi ketepatan dalam memprediksi sebesar 34,3 % dan kesalahan dalam memprediksi sebesar 65,7%.

V. KESIMPULAN

Dalam penelitian ini diusulkan suatu rangkaian proses klasifikasi sel nukleus normal dan abnormal *Pap smear*, dengan menggunakan operasi pengolahan citra yaitu akuisisi citra, preprocessing, analisa tekstur dan terakhir adalah klasifikasi sel nucleus.

Dari hasil yang diperoleh dapat disimpulkan bahwa :

1. Setiap sel *Pap smear* yang telah diproses memiliki berbagai bentuk dan tekstur yang beragam.
2. Tekstur dalam sel akan semakin jelas terlihat terutama pada saat pengujian kelas abnormal, tekstur sel abnormal lebih banyak terlihat dibandingkan dengan tekstur sel normal.
3. Dari hasil perhitungan akurasi kalsifikasi normal dan abnormal, sistem yang dibuat mempunyai akurasi ketepatan dalam memprediksi sebesar 73,6 % dan kesalahan dalam memprediksi sebesar 26,4 %.
4. Untuk hasil perhitungan akurasi klasifikasi kelas sistem yang dibuat mempunyai akurasi ketepatan dalam memprediksi sebesar 34,3 % dan kesalahan dalam memprediksi sebesar 65,7 %.

UCAPAN TERIMA KASIH

Penelitian ini menggunakan data dari Pap smear Benchmark Data For Pattern Classification J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, University Hospital Dept. of Pathology Harlev Ringvej 75, DK-2730 Harlev Denmark.

Denmark:

<http://labs.fme.aegean.gr/decision/downloads/> (25 Desember 2012).

DAFTAR PUSTAKA

- WHO (2013). WHO Guidance note. Number of pages 12 Publication 2013. From <http://www.who.int/2013/01/19/reproductivehealth/publications/cancers/9789241505147/en/index.html>.
- Dalimartha, S. (2004). Deteksi Dini Kanker & Simplisia Antikanker. Jakarta: Penebar Swadaya Jakarta.
- Gonzalez, R.C., R.E. Woods., & Eddins, S.L. (2003). *Digital Image Processing Using MATLAB*, 11-12
- Haralick, R.M., Shanmugan, K., & Dinstein, I. (2003). *Textural Features for Image Classification*, *IEEE Transactions on Systems, Man, and Cybernetics*, 610-621.
- Indriyani, C., & Riana, D. (2010). *Prediction Image Pap Smear Web Based With Decision Tree*. STMIK Nusa Mandiri , 1-5.
- Jantzen, J., Norup, G.J., Dounias., & Bjerregaard, B., (2005). Pap-smear Benchmark Data For Pattern Classification, Technical University of Denmark, 1-20.
- Muhimmah, I., Anwariyah, K., & Indrayanti. (2012). *Extraction and Selection Features of Cervical Cell Types in Pap Smear Digital Images*. *Wise Health ITB*, 1-7.
- Mathworks*. (2012). from Matrix Laboratory: <http://www.mathworks.com/2012/12/25>.
- Martin, E. (2003). *Pap-Smear Classification*. from Technical University of

- , 1-7.
- Novitasari. (2010). Analisis Identifikasi Serviks Normal dan Abnormal Berdasarkan Filter Gabor dan Ekstraksi Ciri Tekstur Statistik. Universitas Gunadarma , 1-7.
- Prasetyo, E. (2011). Pengolahan Citra Digital Dan Aplikasinya Menggunakan Matlab. 1-2.
- Pratama, G., Riana, D., & Hasanudin. (2012). *Pap Smear Nucleus Texture Analysis*. ITB , 1-4.
- Riana, D., widyanto, D. H., & Mengko, T. L. (2012). Perbandingan Segmentasi Luas Nukleus Sel Normal dan Abnormal Pap smear Menggunakan Operasi Kanal Warna dengan Deteksi Tepi Canny dan Rekontruksi Morphologi. *Wise health ITB* , 1-2.
- Selinger, S. (2010). Image Processing and Texture Analysis. *Dennis GaborCollege*
- Sitanggang, G., Carolita, I., & Trisasongko, H. B. (2010). Aplikasi Teknik dan Metode Fusi Data Optik ETM-Plus Landsat dan Sar Radarsat Untuk Ekstraksi Informasi Geologi Pertambangan Batu Bara. *Peneliti Pusbangja-Lapan dan Peneliti IPB* , 1-20.
- Suprpto. (2010). Penggunaan Pengolahan Citra Digital Pada Pemeriksaan Pap Smear Dalam Pendeteksian Kanker Serviks. Universitas Brawijaya , 1-10.
- Sugiyono. (2011). Metode Penelitian Kuantitatif dan Kualitatif dan R&D. Bandung: CV Alfabeta.
- Zuiderveld, K. (2000). Contrast Limited Adaptive Histogram Equalization. *Graphic Gems IV. San Diego: Academic Press Professional* , 474–485.