

Ekstraksi dan Klasifikasi Tekstur Citra Sel Nukleus *Pap Smear*

Dwiza Riana¹, Dwi H. Widyantoro², Tati Latifah R. Mengko³

^{1,2,3} *Teknik Elektro dan Informatika, Institut Teknologi Bandung
Jl. Ganesha 10, Bandung 40132, Indonesia*

¹*dwiza_riana@yahoo.com*

²*dwi@stie.itb.ac.id*

³*tmengko@itb.ac.id*

Abstrak— This paper presents a texture analysis and comparison of classification of cell nucleus images. Texture analysis will be focused on the nuclei of Image Pap smear cell. The method of analysis texture is the statistical second order of Grey Level Co-occurrence Matrix (GLCM). There are five parameter that will be extracted, viz. contrast, correlation, energy, homogeneity and entropy. The image nuclei used in this work are cropped images from Herlev data bank. The images from 917 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) Dysplasia, Moderate Dysplasia, Severe Dysplasia and Carcinoma In Situ. The process of texture analysis developed using grayscale 8 bit level. The preprocessing of images will be conducted before texture analysis in order to improve contrast in nuclei. Based on the numerical result of all parameter, class normal and abnormal of Pap smear image have slightly different properties for correlation, energy, homogeneity and entropy. Originally, there have been 18 fetures of texture which were created to classify into two classes by decision tree classifier, ie normal and abnormal cell. The experimental study shows that in two-class classification, normal and abnormal based on the texture features and using the Decision Tree learning algorithm (J48) classifiers with the Weka Correctly Classification Instances (CCI) and Kappa Coefficient classification performance measures, the Decision Tree learning algorithm (J48) classifier performs the best with the CCI of 73.8277% and the Kappa Coefficient of 0.2785.

Keywords — Pap smear images, nucleus, texture, GLCM, classification, decision tree, cervical cancer

I. PENDAHULUAN

Beban kanker serviks di dunia setiap dua menit seorang wanita meninggal dunia dan di Indonesia setiap satu jam seorang wanita meninggal dunia karena kanker serviks [1]. Kendala sarana dan sumber daya manusia terutama karena sedikitnya ahli patologi menyebabkan upaya deteksi dini kanker serviks cukup sulit dilakukan. Kendala lain adalah kurangnya kesadaran akan penyakit dan akses ke layanan kesehatan. Pencegahan kanker serviks sejak dini dapat dilakukan dengan pemeriksaan rutin Pap smear. Sampai saat ini tahap kunci dalam deteksi otomatis dini kanker serviks adalah akurasi identifikasi sel nukleus [2]. Penelitian tentang nukleus pada citra sel Pap smear banyak dilakukan pada citra berskala abu-abu (*grayscale*) [3]-[10], terutama untuk proses segmentasi nukleus. Penelitian tentang ekstraksi fitur sel Pap smear telah dilakukan sebelumnya. Mustafa N. et al [11] melakukan tentang penelitian ekstraksi fitur yang terdiri dari *perimeter*, *red*, *green*, *blue*, *intensity1*, *intensity2*, dan *saturation*. Selanjutnya membuat klasifikasi dari fitur-fitur tersebut ke dalam tiga kelas sel serviks, yaitu sel serviks normal, LSIL sel, dan HSIL sel. Ekstraksi terhadap nukleus juga telah dilakukan oleh

Plissiti, dkk [12] yang mengidentifikasi tiga fitur yaitu *texture*, *shape* dan *intensity* untuk mengklasifikasi sel nukleus. Ekstraksi tekstur juga telah dilakukan oleh Pratama dkk [13], dengan menggunakan 280 citra sel tunggal, hasil penelitiannya mendapatkan nilai matrik dari setiap testur untuk tujuh kelas dan belum dilakukan klasifikasi [13].

Penelitian tentang klasifikasi sel tunggal *Pap smear* telah banyak dilakukan Amalia dkk. [14] menggunakan fitur kuantitatif dan klasifikasi *Adaptive Network-based Fuzzy Inference System* (ANFIS), selain itu Giri dkk. [15] mengklasifikasi sebagian fitur dengan *association rules*, untuk mengklasifikasi citra sel *Pap smear*. Hasil penelitiannya menunjukkan bahwa untuk hasil pada kasus 7 kelas, akurasi yang dicapai baru mendekati 40% benar.

Dwiza et al. [16] menggunakan 20 fitur kuantitatif dengan klasifikasi *Naïve Bayes (NB)*, *Multi Layer Perceptrons (MLPs)*, *Instance-based Learning algorithm (IBL)*, *Decision Tree learning algorithm (J48)*, dan *Repeated Incremental Pruning to Produce Error Reduction (Jrip)* untuk mengklasifikasi citra sel. Hasil yang diperoleh untuk kasus klasifikasi dua kelas normal dan abnormal, klasifikasi dengan

Decision Tree learning algorithm (J48) menunjukkan performa yang terbaik.

Tujuan penelitian ini melakukan ekstraksi fitur nukleus terhadap 917 sel tunggal terutama untuk nilai statistik tekstur pada nukleus yang terdiri dari *contrast, correlation, energy, homogeneity* dan *entropy* [13]. Kelima fitur ini akan membentuk 18 nilai tekstur sel Pap smear yang selanjutnya akan digunakan untuk klasifikasi sel ke dalam sel normal atau abnormal dengan menggunakan *Decision Tree learning algorithm (J48)*[16]. Ekstraksi fitur tekstur ini akan melengkapi data 20 fitur yang telah dimiliki oleh 917 citra ini, sehingga menjadi 38 fitur.

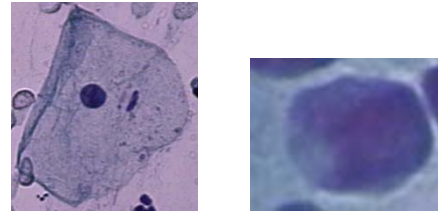
Paper ini terbagi dalam beberapa bagian. Bagian 2 membahas tentang kanker serviks, bagian 3 membahas tentang material yang digunakan dalam penelitian ini, bagian 4 akan membahas tentang metode yang diusulkan dalam ekstraksi tekstur dan klasifikasi, dan bagian 5 menjelaskan tentang pembahasan dan hasil penelitian, selanjutnya ditutup dengan kesimpulan dan rencana penelitian lanjutan.

II. KANKER SERVIKS

Pemeriksaan kanker serviks banyak dilakukan di laboratorium patologi. Jenis pemeriksaan bisa berbentuk konvensional atau *Liquid Base Preparation (LBP)* metode yang berbasis *liquid*. Biasanya hasil LBP lebih baik dari metode konvensional karena dapat mengurangi tampilan darah, lendir, bakteri, jamur, dan nanah yang kadang mempersulit pembacaan di slide konvensional [17].

Sampai saat ini pemeriksaan patologi masih merupakan alat diagnostik terpenting yang harus dilakukan. Pemeriksaan patologi adalah pemeriksaan sampel kecil sel di bawah mikroskop untuk menentukan apakah terdapat kanker dengan melihat abnormalitasnya yaitu membandingkan sel yang diamati dengan sel yang sehat [18].

Kanker serviks adalah penyakit ganas yang berkembang pada sel-sel serviks atau leher rahim. Sel-sel ini tidak tiba-tiba berubah menjadi kanker. Sebaliknya, sel-sel normal dari leher rahim secara bertahap berkembang menjadi prakanker yang kemudian berubah menjadi kanker. Sel kanker menunjukkan peningkatan daerah inti atau nukleus bila dibandingkan dengan sel normal. Fitur yang terkait dengan karakteristik nukleus dapat digunakan untuk melakukan klasifikasi sel-sel serviks normal atau abnormal. Gambar (1a) dan (1b) menunjukkan sel normal dan sel yang abnormal [19]. Sel normal memiliki luas inti yang lebih kecil dan daerah sitoplasma yang sangat besar sedangkan sel yang abnormal sebaliknya terjadi peningkatan luas nukleus tetapi menyusut pada daerah sitoplasma [20].



Gambar (1a) Sel Normal Gambar (1b) Sel Abnormal

III. MATERIAL

Sebanyak 917 citra sel tunggal yang berasal dari data Herlev [21] digunakan dalam penelitian ini. Data citra sel tunggal kanker serviks tersebut telah diklasifikasi dalam tujuh kelas oleh *cyto-technicians* dan dokter berdasarkan pengukuran dan konfirmasi klinikal [19].

Kelas data terdiri dua kelompok yaitu kategori sel normal yaitu *Normal Superficial (NS)*, *Normal Intermediate (NI)*, dan *Normal Columnar (NC)*. Sisanya empat kelas yaitu *Mild (Light) Dysplasia (MLD)*, *Severe Dysplasia (SD)*, *Moderate Dysplasia (MD)*, dan *Carcinoma In Situ (CIS)* adalah sel abnormal [19].

917 citra ini telah memiliki 20 nilai fitur [21] yaitu luas wilayah nukleus atau (*Nucleus Area* atau *Kerne_A*), luas wilayah sitoplasma (*Cytoplasm Area* atau *Cyto_A*). Fitur berikutnya adalah perbandingan luas wilayah nukleus dan sitoplasma (*N/C ratio* atau *K/C*) merupakan perbandingan antara luas wilayah nukleus dengan luas wilayah sitoplasma. Dalam hal ini sebagai pembilang adalah luas wilayah nukleus, sedangkan sebagai penyebutnya adalah luas wilayah sel (sitoplasma ditambah nukleus). Derajat intensitas cahaya nukleus (*Nucleus Brightness* atau *Kerne_Ycol*) dihitung dari nilai rata-rata keterangan atau intensitas cahaya yang dimiliki oleh area nukleus, dapat digunakan fungsi panjang gelombang cahaya warna.

TABEL I
DATA HERLEV SEL CITRA PAP SMEAR

No	Nama Kelas	Jumlah Data
1	Normal Superficial (NS)	74
2	Normal Intermediate (NI)	70
3	Normal Columnar (NC)	98
4	Mild (Light) Dysplasia (MLD)	182
5	Severe Dysplasia (SD)	146
6	Moderate Dysplasia (MD)	197
7	Carcinoma In Situ (CIS)	150
	Total Data	917

Fitur yang lain adalah derajat intensitas cahaya sitoplasma (*Cytoplasm Brightness* atau *Cyto_Ycol*) dihitung dari nilai rata-rata keterangan (intensitas cahaya) yang dimiliki oleh area sitoplasma. Diameter

terpendek nukleus (*Nucleus Shortest Diameter* atau *KerneShort*) merupakan diameter terbesar lingkaran yang dapat dibentuk di dalam area nukleus, dan dapat diukur sebagai jarak terpendek yang menghubungkan antara dua pixel pada batas (*border*) nukleus.

Diameter terpanjang nukleus (*Nucleus Longest Diameter* atau *KerneLong*) merupakan diameter terkecil dari sebuah lingkaran yang dapat dibentuk untuk mengelilingi sebuah nukleus, dan dapat diukur sebagai jarak terpanjang diantara dua pixel pada area batas (*border/lingkar luar*) nukleus. Pemuluran nukleus (*Nucleus Elongation* atau *KerneElong*) merupakan perbandingan antara diameter terpendek dengan diameter terpanjang nukleus. Kebundaran nukleus (*Nucleus Roundness* atau *KerneRund*) merupakan perbandingan antara luas wilayah nukleus sesungguhnya dengan luas wilayah lingkaran yang dibentuk oleh diameter terpanjang nukleus [22].

Diameter terpendek sitoplasma (*Cytoplasm Shortest Diameter* atau *CytoShort*) merupakan diameter terbesar lingkaran yang dapat dibentuk di dalam sebuah sitoplasma, dan dapat diukur sebagai jarak terpendek yang menghubungkan antara dua pixel pada batas sebuah sitoplasma. Diameter terpanjang sitoplasma (*Cytoplasm Longest Diameter* atau *CytoLong*) merupakan diameter terkecil sebuah lingkaran yang dapat dibentuk untuk mengelilingi sebuah sitoplasma, dan dapat diukur sebagai jarak terpanjang yang menghubungkan antara dua pixel pada batas sitoplasma [22].

Pemuluran sitoplasma (*Cytoplasm Elongation* atau *CytoElong*) merupakan perbandingan antara diameter terpendek dengan diameter terpanjang sitoplasma. Kebundaran sitoplasma (*Cytoplasm Roundness* atau *CytoRund*) merupakan perbandingan antara luas wilayah sitoplasma sesungguhnya dengan luas wilayah lingkaran yang dibentuk oleh diameter terpanjang sitoplasma [22].

Batas nukleus (*Nucleus Perimeter* atau *KernePeri*) merupakan panjang batas (*border*) nukleus, atau singkatnya adalah keliling nukleus. Batas sitoplasma (*Cytoplasm Perimeter* atau *CytoPeri*) merupakan panjang batas sitoplasma (keliling sitoplasma). Posisi nukleus (*Nucleus Realtive Position* atau *KernePos*) menjelaskan seberapa baik nukleus berada di tengah sitoplasma, yang dapat dihitung dengan mencari jarak antara pusat nukleus dan pusat sitoplasma. Maksimum di nukleus (*Nucleus Maximum* atau *KerneMax*) merupakan jumlah pixel berlabel nukleus yang berada pada jendela ketetangaan ukuran 3X3 yang berada pada jendela ketetangaan ukuran 3X3 yang berisi pixel nukleus terbanyak. Nilai ini dapat dihitung dengan menentukan terlebih dahulu jumlah pixel nukleus terbanyak yang ada pada sebuah jendela ketetangaan 3X3, contohnya misalkan 6 pixel. Selanjutnya ditentukan berapa banyak jendela

ketetangaan 3X3 yang berisi pixel nukleus terbanyak tersebut, katakan saja ada 10 jendela. Maka nilai Nmax adalah 6x10 yaitu 60 pixel. Minimum di nukleus (*Nucleus Minimum* atau *KerneMin*) merupakan jumlah pixel berlabel nukleus yang berada pada jendela ketetangaan ukuran 3X3 yang berisi pixel nukleus tersedikit. Maksimum di sitoplasma (*Cytoplasm Maximum* atau *CytoMax*) merupakan jumlah pixel berlabel sitoplasma yang berada pada jendela ketetangaan ukuran 3X3 yang berisi pixel sitoplasma terbanyak. Sedangkan fitur terakhir adalah minimum di sitoplasma (*Cytoplasm Minimum* atau *CytoMin*) merupakan jumlah pixel berlabel sitoplasma yang berada pada jendela ketetangaan ukuran 3X3 yang berisi pixel sitoplasma tersedikit [22].

TABEL II
NILAI FITUR SEL NORMAL CITRA PAP SMEAR [21]

Fitur	Normal Superficial		Normal Intermediate		Normal Columnar	
	N_Min	N_Max	N_Min	N_Max	N_Min	N_Max
Kerne_A	316.875	1409.5	568.625	2943.25	673.625	4418.75
Cyto_A	12631.3	127313.75	12507.875	84196	563.625	9610.375
KC	0.00399	1	0.012662859	0.095717393	0.153379368	0.66789423
Kerne_Ycol	17.9139	97.64502165	34.08161709	150.6117928	61.42908654	156.47449
Cyto_Ycol	94.5662	207.7216542	85.50288666	217.3171421	83.82801556	225.759934
KerneShort	13.9739	40.54717639	20.16644984	56.69225592	19.25342658	63.7286268
KerneLong	19.4165	47.16990566	28.28427125	67.41661516	31.57530681	105.233075
KerneElong	0.32819	1.192307692	0.450765864	1.078913325	0.323980143	1.17997097
KerneRund	0.33613	1.135614846	0.430449693	1.067202539	0.281106173	1.05079521
CytoShort	123.833	418.4945404	98.75334845	362.8846717	29.45784971	131.901185
CytoLong	162.791	509.5144748	180.2470527	452.049776	58.18075283	211.021326
CytoElong	0.44418	1.215870819	0.420041204	1.124905157	0.255908321	1.07566766
CytoRund	0.37295	1.071125704	0.420041204	0.871269316	0.095552737	0.62141206
KernePeri	60.75	131.375	0.420041204	196.625	96.875	280
CytoPeri	458.5	1541.75	0.420041204	1408.625	166.375	614.375
KernePos	0.02019	0.494316097	0.420041204	0.422074059	0.036057994	0.87800669
KerneMax	19	64	0.420041204	99	33	171
KerneMin	13	54	0.420041204	80	24	131
CytoMax	398	3906	0.420041204	3255	14	275
CytoMin	444	4000	0.420041204	3311	26	296

Tabel II dan Tabel III memuat nilai-nilai minimum dan minimum untuk kedua puluh fitur untuk citra pada masing-masing kelas. Sehingga dapat memberikan gambaran dari kondisi fitur sel nukleus dan sitoplasma dalam setiap kelas. Nilai ini dirangkum dari nilai 917 citra [21].

Selain 20 fitur ini perlu diekstraksi fitur-fitur lain yang berkaitan dengan citra sel *Pap smear* sehingga bisa didapatkan lebih banyak nilai-nilai fitur dengan harapan dapat lebih memudahkan identifikasi citra sel nukleus. Selain dapat digunakan oleh peneliti-peneliti lain dalam melakukan penelitian untuk mendapatkan klasifikasi terbaik untuk citra sel *Pap smear*.

TABEL III
NILAI FITUR SEL ABNORMAL CITRA PAP SMEAR [21]

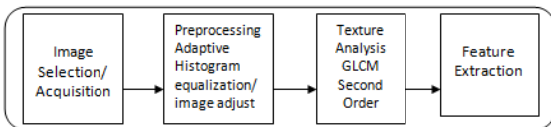
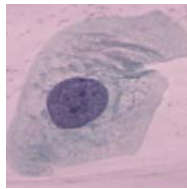
Fitur	Mild (Light) Dysplasia		Severe Dysplasia		Moderate Dysplasia		Carcinoma In Situ	
	N_Min	N_Max	N_Min	N_Max	N_Min	N_Max	N_Min	N_Max
Kerne_A	1121,375	9602,625	1312,125	9800	757,5	10794,625	1160,5	10271,5
Cyto_A	2612,875	62517,5	1030,375	32476,875	594,25	14108,75	467,875	10863,375
KC	0,099453935	0,621777309	0,143762206	0,758475197	0,179315105	0,843201884	0,230741315	0,855497094
Kerne_Ycol	60,17977528	158,8094849	35,51249624	135,4247625	53,25425623	174,9946509	61,98041237	141,5622327
Cyto_Ycol	92,72363926	216,4864702	87,23931367	209,4001809	69,88970174	230,4863985	77,57660338	210,2126697
KerneShort	33,57871547	112,6871807	31,07455051	112,7230534	23,76542375	107,7807264	28,51602835	101,5414105
KerneLong	37,94733192	158,8238017	46,32493929	140,7302384	31,95309062	130,2459212	42,01190308	145,2790418
KerneLong	0,371631553	1,173975215	0,430394602	1,124137931	0,307805597	1,080281426	0,34884765	1,12539827
KerneRand	0,308903207	1,070086902	0,416039473	1,074211286	0,270610283	0,961223132	0,298626991	0,943664937
CytoShort	60,28946156	264,9734387	54,45838612	226,5705541	33,47518855	155,0844643	33,20788364	129,8922927
CytoLong	82,97590012	571,9090837	71,56814934	326,741488	44,64302857	263,3040068	57,00877126	209,3442142
CytoLong	0,322517604	1,191359773	0,33018457	1,177704353	0,295123447	1,100363156	0,210975471	1,162030457
CytoRand	0,190327561	0,689663554	0,12464743	0,807536038	0,074107658	0,742954507	0,06699333	0,633934006
KernePeri	126,375	404,375	131,375	402,375	99,25	384,25	121,125	411,375
CytoPeri	239,125	1718	219,375	935	136,125	666,625	160,25	517,5
KernePos	0,011505825	0,617291901	0,003363582	0,825646304	0,007566562	0,903233198	0,006449227	0,655703669
KerneMax	58	415	59	430	37	425	41	282
KerneMin	25	399	40	432	23	381	25	264
CytoMax	79	2511	43	783	20	357	22	311
CytoMin	81	2472	51	878	21	357	23	330

IV. METODE PENELITIAN

Metode yang diusulkan untuk ekstraksi tekstur terdiri dari proses manual *cropping* nukleus yang bertujuan mengambil citra sel nukleus. Tahapan pertama dalam pendekatan metode penelitian yang diusulkan ekstraksi fitur statistik tekstur citra sel nukleus *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. 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 (1).

$$Gray = 0.2989 * R + 0.5870 * G + 0.1140 * B$$

(1)

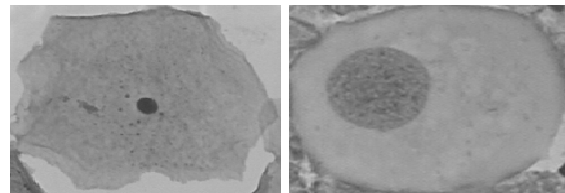


Gambar 2. Metode Penelitian

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 5 parameter yaitu *Entropy*, *Contrast*, *Correlation*, *Energy*, *Homogeneity*, dan *Deviation* Tahap terakhir adalah klasifikasi, pada tahap ini nilai matrik dari

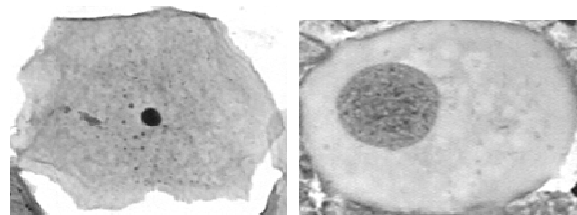
setiap parameter akan diproses untuk proses klasifikasi sel normal dan abnormal. Gambar 2 memperlihatkan alur penelitian yang dilakukan.

Grayscale pada citra sel *Pap smear* memiliki nilai kontras yang rendah, sehingga menyebabkan kesulitan dalam analisis tekstur, gambar 3 menunjukkan hasil citra *grayscale* untuk citra normal dan abnormal. *Preprocessing* 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)*.



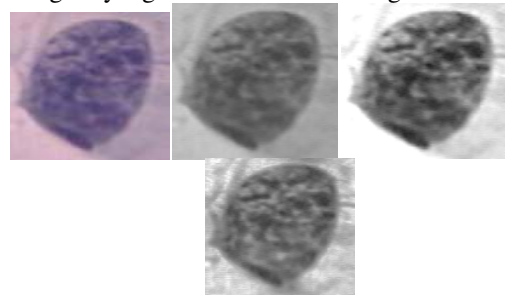
Gambar 3. Citra Hasil Proses *GrayScale*

Image adjustment (*imadjust*) merupakan transformasi intensitas *grayscale* dari sebuah citra. Gambar 4 adalah hasil proses ini untuk citra normal dan abnormal.



Gambar 4. Citra Hasil Proses *Adjustment*

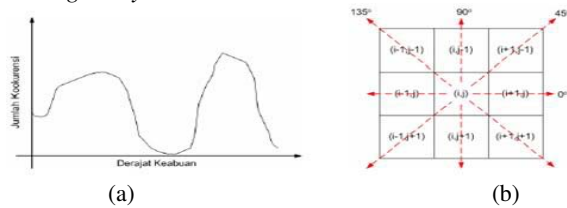
Contrast Limited Adaptive Histogram (CLAHE) adalah teknik perbaikan citra yang digunakan untuk memperbaiki kontras pada citra. CLAHE beroperasi pada region yang kecil dan disebut dengan *tile*.



Gambar 5. Citra Hasil Proses Citra asli Abnormal – Citra CLAHE

Keuntungan menggunakan CLAHE adalah perhitungan yang sederhana, mudah digunakan dan menghasilkan output yang baik, karena memiliki sedikit noise yang sedikit. Gambar 5 memberikan ilustrasi hasil proses untuk citra sel nukleus abnormal.

Analisis tekstur difokuskan pada citra nukleus sel *Pap smear*, menggunakan metode GLCM (*Gray Level Co-occurrence Matrix*). Lima parameter yang digunakan yaitu: *Entropy*, *Contrast*, *Correlation*, *Energy*, *Homogeneity*, dan *Deviation*. 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°. Ekstraksi tekstur menggunakan ke empat arah ini untuk nilai-nilai *Contrast*, *Correlation*, *Energy*, dan *Homogeneity*.



Gambar 8. Contoh Ilustrasi Ekstraksi Ciri Statistik ,

- (a) Histogram citra sebagai fungsi probabilitas kemunculan nilai intensitas pada citra,
- (b) Hubungan ketetanggaan antar piksel sebagai fungsi orientasi dan jarak spasial

Untuk ekstraksi fitur dari sel nukleus secara lengkap rumus yang digunakan untuk fitur-fitur tersebut dapat diberikan pada persamaan berikut ini:

1. Energi (Energy).

$$f1 = \sum_i \sum_j p_d^2(i, j) \tag{2}$$

2. Entropi (Entropy).

$$f2 = - \sum_i \sum_j 2_d(i, j) \log(2_d(i, j)) \tag{3}$$

3. Kontras (Contrast).

$$f3 = \sum_i \sum_j (i - j)^2 p_d(i, j) \tag{4}$$

4. Homogenitas (Homogeneity).

$$f4 = \sum_i \sum_j \frac{p_d(i, j)}{1 + |i - j|} \tag{5}$$

5. Korelasi (Correlation).

$$f5 = \sum_i \sum_j \frac{(i - \mu_x)(j - \mu_y)p_d(i, j)}{\sigma_x \sigma_y} \tag{6}$$

Hasil dari ekstraksi tekstur 18 fitur akan dijadikan acuan untuk mendapatkan rule klasifikasi normal dan abnormal. Dari hasil penelitian terdahulu [16] bahwa klasifikasi *Decision Tree learning algorithm (J48)* untuk kasus dua kelas normal dan abnormal memberikan nilai *Weka Correctly Classification Instances (CCI)* terbaik, maka pada penelitian ini juga digunakan metode klasifikasi yang sama [23, 24].

J48 sudah dikenal memiliki klasifikasi yang cukup baik dan juga merupakan algoritma untuk membangun suatu pohon keputusan (*decision tree*). Model *decision tree* yang terbentuk terdiri dari beberapa bagian seperti simpul dalam (*inside nodes*) yang merepresentasikan nilai dari suatu atribut, cabang (*branches*), dan simpul daun (*leaf nodes*). Simpul teratas disebut juga simpul akar (*root nodes*) [22].

Algoritma ini secara rekursif akan mengunjungi setiap *decision node* untuk memilih split yang optimal hingga tidak mungkin lagi dilakukan proses split. Fungsi rekursif dalam struktur data dan algoritma adalah fungsi yang memanggil dirinya secara langsung atau lewat fungsi lain. Pada algoritma J48, teknik yang digunakan untuk mengukur split optimal yang dihasilkan oleh proses rekursif yaitu menggunakan metode seleksi atribut berupa perhitungan *information gain* [22].

Dalam algoritma ini terdapat metode seleksi atribut berupa perhitungan *information gain* yang digunakan untuk menentukan atribut yang terpilih untuk pembentukan *decision tree*. Perhitungan *information gain* dilakukan pada *decision node* untuk digunakan dalam pemilihan split yang optimal pada pembentukan sebuah model *decision tree* [22].

V. HASIL DAN PEMBAHASAN

Penelitian telah dilakukan dengan 917 citra sel tunggal data Herlev. Sebanyak 242 citra nukleus adalah kelas normal dan 675 citra nukleus kelas abnormal. Seluruh 18 fitur tekstur yang dihasilkan merupakan hasil proses metode penelitian yang diusulkan.

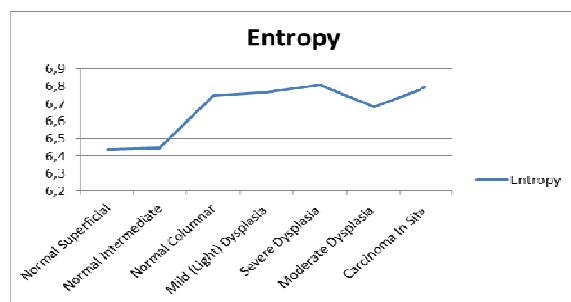
TABEL IV
NILAI EKSTRAKSI TEKSTUR SEL NORMAL CITRA PAP SMEAR

Parameter Tekstur	Normal Superficial		Normal Intermediate		Normal Columnar	
	N_Min	N_Max	N_Min	N_Max	N_Min	N_Max
Entropy	0,865632257	6,435847027	1,230534284	6,445085469	1,557683159	6,743115141
Contrast 0	0,017434379	0,176679842	0,039385236	0,243730338	0,034076433	0,257374205
Contrast 45	0,027351379	0,294216571	0,068496613	0,372731919	0,053374184	0,514779874
Contrast 90	0,020865261	0,251292718	0,0535052	0,300893908	0,045825989	0,43796875
Contrast 135	0,028093711	0,294353295	0,075654636	0,418766016	0,055470451	0,527201258
Correlation 0	0,901832308	0,985529952	0,903531658	0,985220631	0,912688295	0,9861841
Correlation 45	0,848629089	0,976173004	0,846293977	0,975932283	0,870548812	0,975027012
Correlation 90	0,872178737	0,980642346	0,867061131	0,982282234	0,895959505	0,987766774
Correlation 135	0,855940186	0,976324527	0,842210423	0,975170125	0,881932581	0,980680428
Energy 0	0,149720851	0,903861599	0,121518262	0,806484985	0,101438181	0,710220711
Energy 45	0,129056738	0,899354056	0,10398149	0,803341631	0,074671641	0,699051101
Energy 90	0,134876677	0,902265452	0,111942779	0,805973117	0,079801367	0,704334273
Energy 135	0,131087869	0,899386979	0,10106487	0,802874425	0,074552777	0,69904278
Homogeneity 0	0,91468811	0,991823412	0,890581052	0,982524565	0,876325429	0,983174098
Homogeneity 45	0,866533361	0,988739392	0,846524418	0,975847909	0,805398323	0,975575752
Homogeneity 90	0,87498623	0,990691076	0,86009255	0,979454567	0,82484375	0,97731294
Homogeneity 135	0,861788162	0,988479987	0,843191782	0,974406531	0,803533138	0,972587277
Deviation	21,54590311	69,716783	26,4441645	80,45500938	17,62692265	86,50344678

TABEL IV
NILAI EKSTRAKSI TEKSTUR SEL ABNORMAL CITRA PAP SMEAR

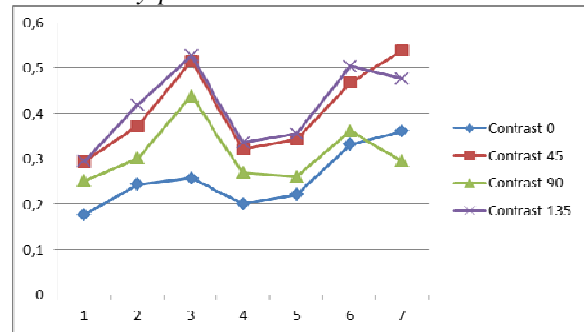
Parameter Tekstur	Mild (Light) Dysplasia		Severe Dysplasia		Moderate Dysplasia		Carcinoma In Situ	
	N_Min	N_Max	N_Min	N_Max	N_Min	N_Max	N_Min	N_Max
Entropy	1,699404487	6,763040028	1,143782602	6,804494851	1,709473381	6,682587074	3,349137827	6,7
Contrast 0	0,032145224	0,201234508	0,017755839	0,221456693	0,035443205	0,331272894	0,065513547	0,3
Contrast 45	0,051693212	0,321142979	0,030417477	0,34303927	0,050426326	0,467952582	0,101570357	0,5
Contrast 90	0,040132389	0,268759248	0,024257517	0,260526316	0,044113757	0,361507937	0,090399386	0,2
Contrast 135	0,051816917	0,335515628	0,026492641	0,355113872	0,061417533	0,503787879	0,105979732	0,4
Correlation 0	0,862535928	0,9884855	0,940527635	0,991320656	0,898235233	0,990090339	0,870133743	0,9
Correlation 45	0,792326581	0,98279096	0,904484717	0,987716393	0,862441815	0,984535792	0,812124071	0,9
Correlation 90	0,810057053	0,983415647	0,922011785	0,989839197	0,875098682	0,987116627	0,835375789	0,9
Correlation 135	0,805973653	0,981167209	0,906988108	0,987487763	0,83475841	0,983319401	0,803474884	0,9
Energy 0	0,109699146	0,74010637	0,095025622	0,838865718	0,102752391	0,793073952	0,101599146	0,4
Energy 45	0,095669664	0,733468269	0,080325198	0,835269977	0,088003502	0,782706626	0,085698498	0,4
Energy 90	0,100180863	0,737402613	0,090789871	0,837708574	0,093384389	0,787814853	0,096082777	0,4
Energy 135	0,091918405	0,733740749	0,080425379	0,836863885	0,08199693	0,78025259	0,088820406	0,4
Homogeneity 0	0,905362396	0,984502694	0,891391581	0,991122081	0,858461423	0,983845951	0,860054628	0,9
Homogeneity 45	0,869692982	0,977202206	0,848454746	0,985704588	0,829124579	0,976013634	0,825165497	0,9
Homogeneity 90	0,889928699	0,98127694	0,875974659	0,988215581	0,854633085	0,979409171	0,870909867	0,9
Homogeneity 135	0,862612045	0,977458636	0,849126984	0,987123947	0,798926768	0,974631406	0,828068824	0,9
Deviation	20,34640867	81,31345878	27,31027781	94,59131679	15,44609113	89,88716272	19,92003742	89,

Hasil ekstraksi tekstur *entropy* nilai meningkat seiring dengan keabnormalan sel. Nilai entropy tertinggi pada kelas *Severe Dysplasia* dan nilai entropy terendah pada dua kelas normal yaitu *Normal Superficial* dan *Normal Intermediate* (Gambar.6).



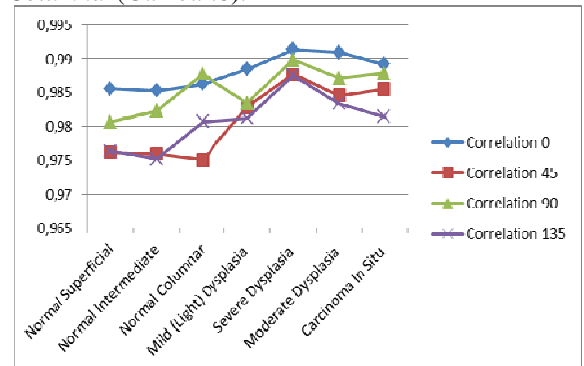
Gambar 6. Grafik Nilai Ekstraksi Tekstur Fitur *Entropy*

Grafik nilai fitur *contrast* untuk (0° , 45° , 90° , dan 135°) pada Gbr. 7 terlihat memiliki pola yang sama. Kecuali untuk kelas yang paling tidak normal *Carcinoma In Situ* memiliki dua pola (0° dan 45°) cenderung sama dan berbeda dengan 90° , dan 135° . Dari semua arah nilai *contrast* memiliki nilai yang konsisten tinggi pada kelas *Normal Columnar* dan *Moderate Dysplasia*.



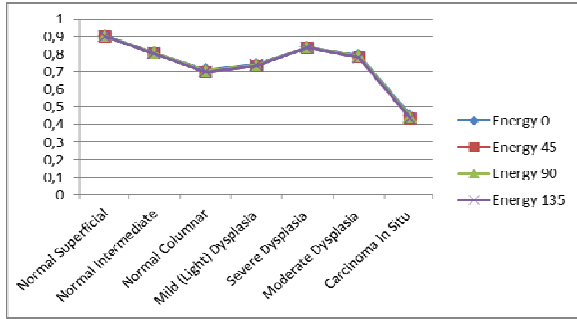
Gambar 7. Grafik Nilai Ekstraksi Tekstur Fitur *Contrast*

Grafik nilai fitur *Correlation* memiliki nilai yang lebih tinggi pada citra sel abnormal. Ada sedikit perbedaan pola pada kelas ke tiga yaitu *Normal Columnar* (Gambar.8).



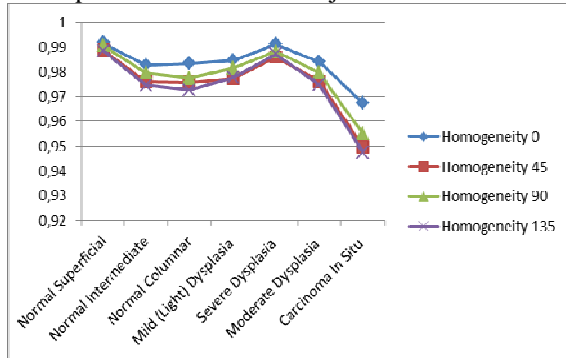
Gambar 8. Grafik Nilai Ekstraksi Tekstur Fitur *Correlation*

Fitur *energy* menunjukkan hasil ekstraksi yang konsisten untuk semua arah sehingga grafik bertumpuk pada nilai yang sama. Dengan kata lain tidak ada perbedaan nilai untuk arah (0° , 45° , 90° , dan 135°). Nilai tertinggi *energy* dimiliki kelas normal (*Normal Superficial*) dan nilai terendah pada kelas *Carcinoma In Situ* yang merupakan kelas abnormal. Jika dilihat sebaran hasil ekstraksi ini fitur *energy* dapat diprediksi sebagai fitur pembeda yang dapat memisahkan kelas dengan baik (Gambar. 9).



Gambar 9. Grafik Nilai Ekstraksi Tekstur Fitur Energy

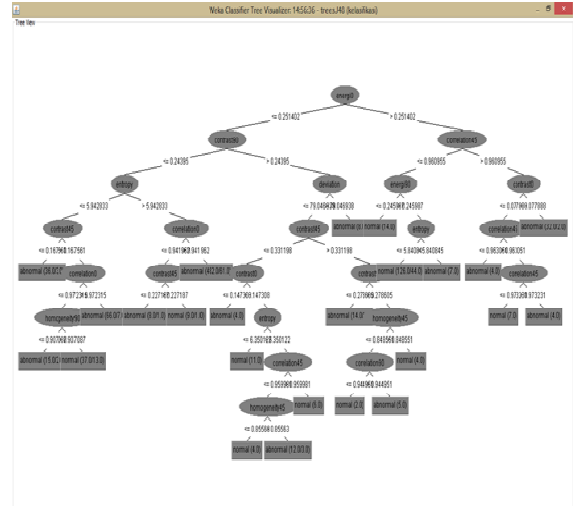
Untuk fitur *homogeneity* memiliki pola nilai yang sama untuk setiap arah. Nilai tertinggi berada pada kelas yang normal dan nilai terendah berada pada sel abnormal. Tetapi pada kelas *abnormal severe dysplasia* terlihat memiliki nilai *homogeneity* cukup tinggi hampir mendekati nilai normal superficial. Grafik pada Gambar.10 menunjukkan kondisi ini.



Gambar 10. Grafik Nilai Ekstraksi Tekstur Fitur Homogeneity

Setelah menganalisa hasil ekstraksi tekstur citra sel nukleus maka proses selanjutnya adalah melakukan klasifikasi terhadap 18 fitur tersebut.

Data yang digunakan adalah semua nilai yang diperoleh untuk 917 citra. Gambar 11 merupakan hasil *decision tree* dari algoritma J48, dengan penggunaan 18 fitur terlihat *tree* yang dihasilkan memiliki *leaf* yang cukup banyak sejumlah 24 dan size *tree* sebesar 47. Dari *decision tree* tersebut diperoleh rule klasifikasi seperti pada Gambar. 13. Rule ini merupakan klasifikasi untuk kasus dua kelas yaitu memisahkan kelas normal dan abnormal.



Gambar 12. Decision Tree

```

energi0 <= 0.251402
|
| contrast90 <= 0.243395
| |
| | entropy <= 5.942833
| | |
| | | contrast45 <= 0.167561: abnormal (36.0/3.0)
| | | contrast45 > 0.167561
| | | |
| | | | correlation0 <= 0.972315
| | | | |
| | | | | homogeneity90 <= 0.907087: abnormal
| | | | | (15.0/2.0)
| | | | | |
| | | | | | homogeneity90 > 0.907087: normal (37.0/13.0)
| | | | | | correlation0 > 0.972315: abnormal (66.0/7.0)
| | | | | entropy > 5.942833
| | | | | |
| | | | | | correlation0 <= 0.941962
| | | | | | |
| | | | | | | contrast45 <= 0.227187: abnormal (8.0/1.0)
| | | | | | | contrast45 > 0.227187: normal (9.0/1.0)
| | | | | | | correlation0 > 0.941962: abnormal (482.0/61.0)
| | | | | contrast90 > 0.243395
| | | | | |
| | | | | | deviation <= 79.046938
| | | | | | |
| | | | | | | contrast45 <= 0.331198
| | | | | | | |
| | | | | | | | contrast0 <= 0.147308: abnormal (4.0)
| | | | | | | | contrast0 > 0.147308
| | | | | | | | |
| | | | | | | | | entropy <= 6.350122: normal (11.0)
| | | | | | | | | entropy > 6.350122
| | | | | | | | | |
| | | | | | | | | | correlation45 <= 0.959981
| | | | | | | | | | |
| | | | | | | | | | | homogeneity45 <= 0.85563: normal (4.0)
| | | | | | | | | | | homogeneity45 > 0.85563: abnormal
| | | | | | | | | | | (12.0/3.0)
| | | | | | | | | | | |
| | | | | | | | | | | | correlation45 > 0.959981: normal (6.0)
| | | | | | | | | | | | contrast45 > 0.331198
| | | | | | | | | | | | |
| | | | | | | | | | | | | contrast90 <= 0.278605: abnormal (14.0/1.0)
| | | | | | | | | | | | | contrast90 > 0.278605
| | | | | | | | | | | | | |
| | | | | | | | | | | | | | homogeneity45 <= 0.848551
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | correlation90 <= 0.944951: normal (2.0)
| | | | | | | | | | | | | | | correlation90 > 0.944951: abnormal (5.0)
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | homogeneity45 > 0.848551: normal (4.0)
| | | | | | | | | | | | | | | | deviation > 79.046938: abnormal (8.0)
energi0 > 0.251402
|
| correlation45 <= 0.960955
| |
| | energi90 <= 0.245987: normal (14.0)
| | energi90 > 0.245987
| | |
| | | entropy <= 5.840845: normal (126.0/44.0)
| | | entropy > 5.840845: abnormal (7.0)
| | correlation45 > 0.960955
| | |
| | | contrast0 <= 0.077888
    
```

```

| | | correlation45 <= 0.963051: abnormal (4.0)
| | | correlation45 > 0.963051
| | | | correlation45 <= 0.973231: normal (7.0)
| | | | correlation45 > 0.973231: abnormal (4.0)
| | | contrast0 > 0.077888: abnormal (32.0/2.0)
    
```

Gambar 13. Rule Hasil Klasifikasi Decision Tree

VI. KESIMPULAN

Dalam penelitian ini diusulkan satu rangkaian proses ekstraksi tekstur citra sel nukleus dan klasifikasi berdasarkan 18 fitur yang diperoleh ke dalam dua kelas sel normal dan abnormal dengan menggunakan *Decision Tree learning algorithm* (J48). *Sel Pap smear* memiliki variasi fitur dari tekstur.

Secara keseluruhan perbandingan hasil ekstraksi 18 tekstur dapat disimpulkan hal-hal sebagai berikut:

1. Untuk fitur *contrast*, *energy* dan *homogeneity* pada arah (0^0 , 45^0 , 90^0 , dan 135^0) cenderung terlihat memiliki pola yang sama. Jadi penggunaan arah tidak membuat perbedaan hasil bagi ketiga fitur ini.
2. Untuk sel abnormal memiliki nilai *correlation* dan *entropy* yang lebih tinggi dibandingkan dengan sel normal. Artinya untuk ketiga fitur ini mengalami kenaikan nilai seiring dengan keabnormalan sel.
3. Untuk sel normal memiliki nilai *energy* dan *homogeneity* tertinggi berada pada kelas yang normal.
4. Hasil klasifikasi untuk kelas normal dan abnormal pada 917 citra dengan menggunakan *Decision Tree learning algorithm* (J48) diperoleh tingkat akurasi Weka CCI (*Correctly Classified Instances*) sebesar 677 atau prosentasi akurasi 73.8277 % dengan *Incorrectly Classified Instances* sebanyak 240 atau sebesar 26.1723% dengan nilai *Kappa statistic* 0.2785.

Penelitian lanjutan dapat dilakukan dengan melakukan pemilihan fitur sebelum dilakukan klasifikasi. Atau melakukan kombinasi dan gabungan dari 38 fitur yang ada. Selain itu dapat dilakukan klasifikasi dengan metode klasifikasi yang lain dengan tujuan untuk meningkatkan performa klasifikasi.

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 Herlev Ringvej 75, DK-2730 Herlev, Denmark.

REFERENSI

- [1] Kusuma, Fitriyadi. "Tes Pap dan Cara Deteksi Dini Kanker Serviks Lainnya," Departemen Obstetri dan Ginekologi Fakultas Kedokteran Universitas Indonesia RSUPN dr. Cipto Mangunkusumo, Prevention and Early Detection of Cervical Cancer (PEACE), Jakarta, 2012.
- [2] Kale, As and Aksoy, Selim, "Segmentation of Cervical Cell Images", International Conference on Pattern Recognition, IEEE, 2010.
- [3] P. Bamford and B. Lovell, "A water immersion algorithm for cytological image segmentation," in Proc. APRS Image Segmentation Workshop, Sydney, Australia, 1996, pp. 75–79.
- [4] P. Bamford and B. Lovell, "Unsupervised cell nucleus segmentation with active contours," Signal Process., vol. 71, no. 2, pp. 203–213, 1998.
- [5] H. S. Wu, J. Barba, and J. Gil, "A parametric fitting algorithm for segmentation of cell images," IEEE Trans. Biomed. Eng., vol. 45, no. 3, pp. 400–407, Mar. 1998.
- [6] A. Garrido and N. P. de la Blanca, "Applying deformable templates for cell image segmentation," Pattern Recognit., vol. 33, no. 5, pp. 821–832, 2000
- [7] N. Lassouaoui and L. Hamami, "Genetic algorithms and multifractal segmentation of cervical cell images," in Proc. 7th Int. Symp. Signal Process. Appl., 2003, vol. 2, pp. 1–4.
- [8] E. Bak, K. Najarian, and J. P. Brockway, "Efficient segmentation framework of cell images in noise environments," in Proc. 26th Int. Conf. IEEE Eng. Med. Biol., Sep., 2004, vol. 1, pp. 1802–1805.
- [9] N. A. Mat Isa, "Automated edge detection technique for Pap smear images using moving K-means clustering and modified seed based region growing algorithm," Int. J. Comput. Internet Manag., vol. 13, no. 3, pp. 45–59, 2005.
- [10] C. H. Lin, Y. K. Chan, and C. C. Chen, "Detection and segmentation of cervical cell cytoplasm and nucleus," Int. J. Imaging Syst. Technol., vol. 19, no. 3, pp. 260–270, 2009.
- [11] N. Mustafa, N.A. Mat Isa, M. Y. Mashor and N. H. Othman, "Capability Of New Features Of Cervical Cells For Cervical Cancer Diagnostic System Using Hierarchical Neural Network" IJSSST, Vol. 9, No. 2, May 2008. Pp. 56-64
- [12] M.E. Plissiti, C. Nikou and A. Charchanti "Combining Shape, Texture And Intensity Features For Cell Nuclei Extraction In Pap Smear Images", Pattern Recognition Letters, vol.32, No 6, pp 838 – 853, 2011.
- [13] Pratama, G.K., Riana, Dwiza, Ekashanti Octorina Dewi, Dyah, Widyantoro, Dwi H and Tati. LM, "Pap smear Nuclei Tekstur Analysis". International Conference on Women's Health in Science & Engineering (WiSE Health), ITB, Bandung. 2012
- [14] Amalia, Ratih. Pap Smear Cell Image Classification Using Adaptive Network-based Fuzzy Inference System (ANFIS) and Biomedical Image Retrieval System. Final Project. Faculty of Computer Science, University of Indonesia, (2007)
- [15] Giri, Endang P. Pap smear Image Classification Based On Association Rules for Biomedical Image Retrieval System. Thesis. Faculty of Computer Science, University of Indonesia, 2008
- [16] Dwiza, R. and Murni, A. Performance Evaluation of Pap Smear Cell Image Classification Using Quantitative and Qualitative Features Based On Multiple Classifier. Conference International ICACSSIS, 2009.
- [17] Purwadi, Sigit. "Indonesian Cervical Cancer Challenge", Divisi Oncologi Dept Obstetrics Gynecology Faculty of Medicine. Universitas Indonesia, Prevention and Early Detection of Cervical Cancer (PEACE) 2012
- [18] Koswara, Teja, "Pathological diagnosis of cervical cancer", Seminar Club Biomedical Engineering, STEI ITB, Bandung, 2012.

- [19] Martin, Erik. Pap-Smear Classification. Technical University of Denmark – DTU.2003.<http://fuzzy.iau.dtu.dk/download/martin2003>
- [20] BustanurRosidi, NorainiJalil, Nur. M. Pista, Lukman H. Ismail, EkoSupriyantoTati L. Mengko “Classification of Cervical Cells Based on Labeled Colour Intensity Distribution” International Journal of Biology and Biomedical Engineering, Issue 4, Volume 5, 2011
- [21] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, Pap-smear Benchmark Data For Pattern Classification, Technical University of Denmark, Denmark, 2005.
- [22] Dwiza, riana. Hierarchical Decision Approach Berdasarkan Importance Performance Analysis Untuk Klasifikasi Citra Tunggal Pap Smear Menggunakan Fitur Kuantitatif Dan Kualitatif. Thesis. Faculty of Computer Science, University of Indonesia, 2010.
- [23] Witten, I.H. and Frank, G. Data Mining: Practical Machine Learning Tools with Java Implementations, Morgan Kaufmann, San Francisco (2000)
- [24] Ryan, Shevaun and Hall, Mark. Practical Data Mining, University of Waikato, August 2006.