

KLASIFIKASI INTI SEL PAP SMEAR BERDASARKAN ANALISIS TEKSTUR MENGUNAKAN *CORRELATION-BASED* FEATURE SELECTION BERBASIS ALGORITMA C4.5

Toni Arifin

Teknik Informatika

Universitas BSI Bandung

Jalan Sekolah Internasional No. 1-6, Bandung 40282

toni.tfn@bsi.ac.id

***Abstract** - Pap Smear is an early examination to diagnose whether there's indication cervical cancer or not, the process of observations were done by observing pap smear cell under the microscope. There's so many research has been done to differentiate between normal and abnormal cell. In this research presents a classification of pap smear cell based on texture analysis. This research is using the Harlev image which amounts to 280 images, 140 images are used as training data and 140 images other are used as testing. On the texture analysis used Gray level Co-occurrence Matrix (GLCM) method with 5 parameters that is correlation, energy, homogeneity and entropy added by counting the value of brightness. For choose which the best attribute used correlation-based feature selection method and than used C45 algorithm for produce classification rule. The result accuracy of the classification normal and abnormal used decision tree C45 is 96,43% and errors in predicting is 3,57%.*

***Keywords** : Classification, Pap Smear cell image, texture analysis, Correlation-based feature selection, C45 algorithm.*

Abstrak - Pap Smear merupakan pemeriksaan dini untuk mendiagnosa apakah ada indikasi kanker serviks atau tidak, proses pengamatan dilakukan dengan mengamati sel pap smear dibawah mikroskop. Banyak penelitian yang telah dilakukan untuk membedakan antara sel normal dan abnormal. Dalam penelitian ini menyajikan klasifikasi inti sel pap smear berdasarkan analisis tektur. Citra yang digunakan dalam penelitian ini adalah citra Harlev yang berjumlah 280 citra, 140 citra digunakan sebagai data training dan 140 citra lain digunakan sebagai testing. Pada analisis tekstur menggunakan metode Gray level Co-occurrence Matrix (GLCM) menggunakan 5 parameter yaitu korelasi, energi, homogenitas dan entropi ditambah dengan menghitung nilai brightness. Untuk memilih mana atribut terbaik digunakan metode correlation-based feature selection lalu digunakan algoritma C45 untuk menghasikan rule klasifikasi. akurasi yang dihasilkan dari klasifikasi normal dan abnormal yang menggunakan decision tree C45 adalah 96,43% dan kesalahan dalam memprediksi adalah 3,57%.

Kata Kunci: Klasifikasi, Citra sel Pap Smear, Analisis tekstur, Correlation-based feature selection, Algoritma C45.

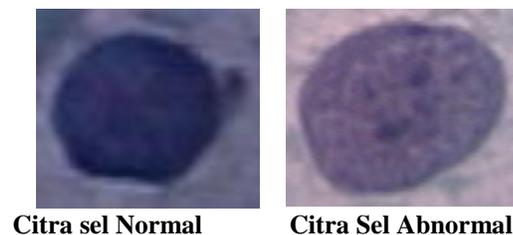
PENDAHULUAN

Kanker serviks atau yang disebut juga sebagai kanker mulut rahim merupakan salah satu penyakit kanker yang paling banyak ditakuti kaum wanita. Berdasarkan data dari WHO dari sekian banyak penderita kanker di Indonesia, penderita kanker serviks mencapai sepertiganya, dan data dari WHO tercatat, setiap tahun ribuan wanita meninggal karena penyakit kanker serviks ini dan merupakan jenis kanker yang menempati peringkat teratas sebagai penyebab kematian wanita di dunia banyak faktor penyebab kanker serviks salah satunya adalah kebiasaan hidup yang kurang baik seperti kebiasaan merokok kurangnya asupan vitamin terutama vitamin C dan vitamin E serta kurangnya asam folat, kebiasaan lainnya yang menyebabkan kanker serviks adalah seringnya melakukan hubungan intim dengan pria yang sering berganti pasangan dan melakukan hubungan intim pada usia dini faktor lain penyebab kanker serviks adalah adanya keturunan kanker. Pemeriksaan rutin dengan Pap Smear dapat membantu mencegah secara dini kanker serviks. Pemeriksaan terhadap sel serviks dilakukan dibawah mikroskop oleh ahli patologi anatomi untuk menyatakan hasil pemeriksaan Pap Smear dari pasien wanita apakah memiliki sel normal atau abnormal. selama ini banyak penelitian yang telah dilakukan untuk mendeteksi sel Pap Smear normal dan abnormal dengan menggunakan citra Sel Pap Smear yang sebelumnya telah di digitalisasi terlebih dahulu sebelum diolah diprogram tapi akurasi masih kurang dari 100%. Tujuan dari penelitian ini adalah memperbaiki tingkat akurasi klasifikasi sel normal dan abnormal, dengan cara pemilihan atribut terbaik lalu dilakukan klasifikasi citra sel dengan menggunakan algoritma decision tree C45.

KAJIAN LITERATUR

Kanker serviks adalah salah satu kanker yang paling banyak menyerang wanita di Indonesia selain kanker payudara. Kanker yang menyerang leher rahim dan juga merupakan kanker nomor 2 terbanyak yang menyerang wanita di dunia. Kanker serviks disebabkan oleh infeksi *Human Papilloma Virus* (HPV). virus ini sebagian besar menular lewat hubungan seksual, namun pada kasus yang jarang, bisa juga

terinfeksi karena pola hidup yang kurang bersih. Gejala dari virus ini disertai keputihan, sedangkan kanker ovarium dan kanker rahim tidak. HPV ada bermacam jenis. HPV tipe 16 dan 18 merupakan penyebab utama pada 70% kasus kanker serviks di dunia. Sedangkan HPV 6 dan 11 adalah penyebab utama kutil kelamin. infeksi HPV seringkali tidak menimbulkan gejala dan baru menyebabkan kanker setelah 10-17 tahun kemudian. Ketika terdeteksi pun seringkali sudah memasuki stadium lanjut yang lebih sulit diobati. Test atau pemeriksaan Pap Smear adalah metode (*screening*) ginekologi, merupakan pemeriksaan leher rahim (serviks) menggunakan alat yang dinamakan *speculum*, dan bisa dilakukan oleh dokter kandungan. Pemeriksaan ini dilakukan untuk mengetahui adanya HPV ataupun sel karsinoma penyebab kanker leher rahim, sejak dini. Meski Pap Smear hanya metode *skrining* yang fungsinya untuk pencegahan kanker serviks, namun metode ini mampu mendeteksi lebih dari 90%. Kanker leher rahim tahap awal masih mungkin untuk disembuhkan. Karakteristik sel dapat digunakan sebagai penanda apakah sel yang diamati merupakan normal atau abnormal. Sel normal memiliki luas nucleus lebih kecil dan luas sitoplasma yang sangat besar sedangkan sel yang abnormal pada nukleus telah meningkat sehingga luas sitoplasma menyusut. Gambar dibawah ini menunjukkan citra sel normal dan abnormal.



Gambar 1
Citra Sel Normal Dan Abnormal

METODE PENELITIAN

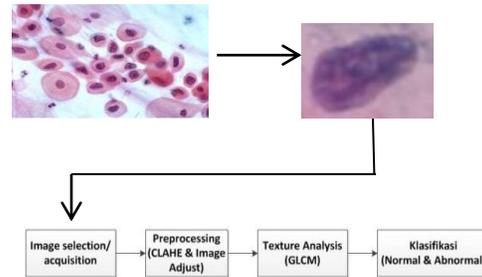
Penelitian ini menggunakan 140 citra dari 280 citra data Harlev. Data Citra sel tunggal kanker serviks tersebut telah diklasifikasikan ke

dalam tujuh kelas *cyto-technicians* berdasarkan pengukuran dan konfirmasi klinikal.

Tabel 1
Data Harlev Citra Sel PAP SMEAR

No	Nama Kelas	Jumlah Data	Jumlah sample
1	Normal Superficial	40	20
2	Normal Intermediate	40	20
3	Normal Columnar	40	20
4	Mild (Light) Dysplasia	40	20
5	Severe Dysplasia	40	20
6	Moderate Dysplasia	40	20
7	Carcinoma In Situ	40	20
Total		280	140

Data citra sel tunggal Pap Smear dari database Harlev dapat diunduh untuk kegiatan penelitian serta dapat digunakan untuk analisis dan validasi. Data citra tersebut sudah dikelompokkan ke dalam 7 kelas yaitu 3 kelas normal, *Normal Superficial*, *Normal Intermediate*, *Normal Columnar* dan 4 kelas abnormal, *Mild (Light) Dysplasia*, *Severe Dysplasia*, *Moderate Dysplasia* dan *Carcinoma In Situ*. Dari 280 citra 140 citra digunakan untuk *training* dan 140 citra digunakan untuk *testing*. Tahapan pertama dalam metode penelitian yang diusulkan untuk proses klasifikasi sel normal dan abnormal Pap Smear adalah mengkonversi citra RGB ke grayscale tanpa merubah kontras pada citra asli pada tahap ini citra 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 matrix yang nantinya akan digunakan untuk klasifikasi pada tahap akhir.

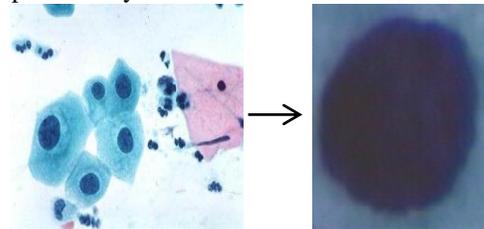


Gambar 2
Metode Penelitian

PEMBAHASAN

Proses Cropping dan Akuisisi Citra

Pada proses pertama dilakukan proses cropping citra untuk mengambil objek dari penelitian yaitu adalah nukleus.



Gambar 3
Proses Cropping Citra

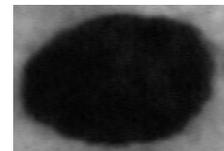
Preprocessing

Grayscale pada citra sel Pap smear memiliki nilai kontras yang rendah, sehingga menyebabkan kesulitan dalam analisis tekstur. Preprocessing 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)*. Dibawah ini adalah proses konversi citra dari citra RGB ke *Grayscale*

Citra sel Normal

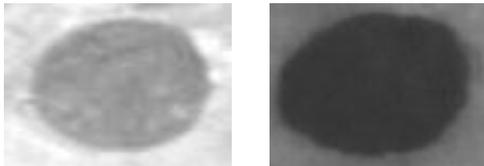


Citra Sel Abnormal



Gambar 4
Citra Normal dan Abnormal Hasil Konversi ke Grayscale

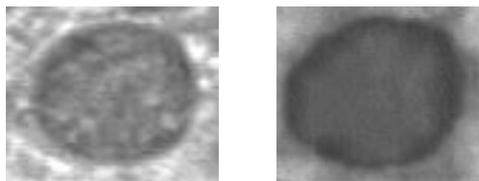
Image adjustment (imadjust) merupakan *basic tool* dari *image processing toolbox* yang disediakan oleh Matlab untuk transformasi intensitas grayscale dari sebuah citra, persamaannya adalah sebagai berikut.



Citra Sel Normal Citra Sel Abnormal

Gambar 5
Citra Normal dan Abnormal Hasil Konversi ke 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.

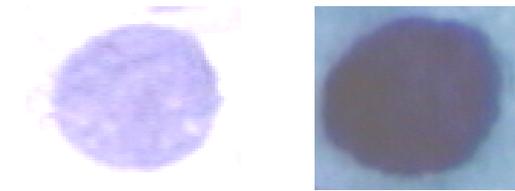


Citra Sel Normal Citra Sel Abnormal

Gambar 6
Citra Normal dan Abnormal Hasil Proses Dengan Teknik CLAHE

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 matriks, nilai matriks ini akan diolah bersama nilai matriks 5 parameter dan kesemua nilai akan digunakan untuk klasifikasi citra.

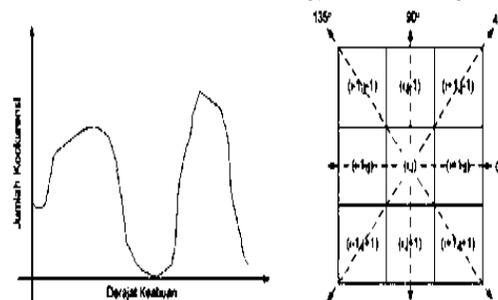


Citra Sel Normal Citra Sel Abnormal

Gambar 7
Citra Normal dan Abnormal Hasil Konversi ke Brightness

Analisis Tekstur

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 9
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

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

1. Energi (*Energy*).

$$f1 = \sum_i \sum_j p_d^2(i, j) \quad (1)$$

2. Entropi (*Entropy*).

$$f2 = - \sum_i \sum_j 2_d(i, j) \log(2_d(i, j)) \quad (2)$$

3. Kontras (*Contrast*).

$$f3 = \sum_i \sum_j (i - j)^2 p_d(i, j) \quad (3)$$

4. Homogenitas (*Homogeneity*).

$$f4 = \sum_i \sum_j \frac{p_d(i, j)}{1 + |i - j|} \quad (4)$$

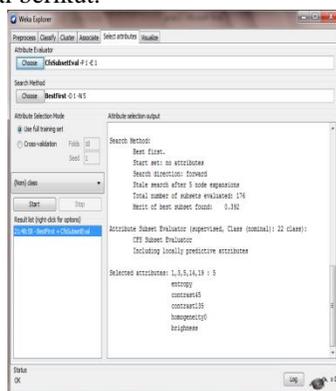
5. Korelasi (*Correlation*).

$$f5 = \sum_i \sum_j \frac{(i - \mu_x)(j - \mu_y)p_d(i, j)}{\sigma_x \sigma_y} \quad (5)$$

Hasil dari ekstraksi tekstur 18 fitur akan dijadikan acuan untuk mendapatkan rule klasifikasi normal dan abnormal, sebelum melakukan proses klasifikasi digunakan metode *correlation-based feature selection* untuk memilih atribut terbaik dari 18 atribut yang ada, setelah itu atribut yang telah terpilih akan digunakan untuk mendapatkan rule klasifikasi dengan menggunakan algoritma C4.5.

Correlation-based Feature Selection (CFS)

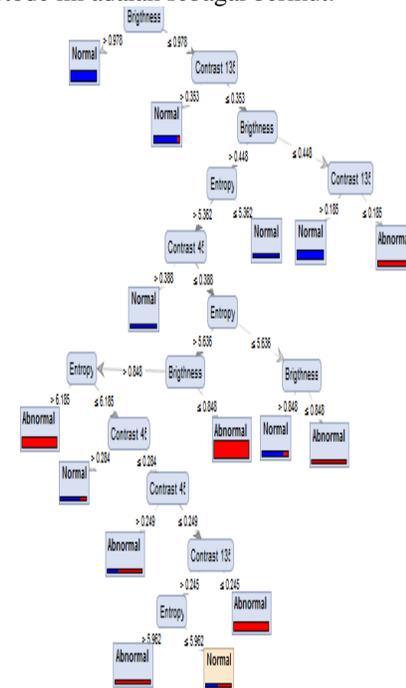
Metode Correlation-based Feature Selection atau yang disingkat CFS merupakan metode seleksi atribut dengan cara menghitung dan membandingkan tingkat korelasi antara masing-masing variable terhadap variable kelasnya dan antara atribut itu sendiri. Dari penelitian ini maka di pilih atribut terbaik yaitu sebagai berikut.



Gambar 10
Atribut terbaik dari 18 Atribut

Algoritma Decesion Tree C4.5

Algoritma C4.5 merupakan salah satu metode klasifikasi menarik yang melibatkan konstruksi pohon keputusan. Koleksi *node* keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar sampai berakhir di *node* daun. Dimulai dari *node* root, ditempatkan dibagian atas dari diagram pohon keputusan, atribut diuji pada *node* keputusan, dengan setiap hasil yang mungkin menghasilkan cabang. Setiap cabang kemudian mengarah ke *node* keputusan yang baik atau ke node daun untuk mengakhiri. Pohon keputusan mirip sebuah struktur pohon dimana terdapat *node* internal (bukan daun) yang mendeskripsikan atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Rule yang dihasilkan dengan metode ini adalah sebagai berikut.



Gambar 11
Decesion Tree

Tree

```

Brightnes > 0.378: Normal (Normal=19, Abnormal=0)
Brightnes < 0.378
| Contrast 135 > 0.353: Normal (Normal=10, Abnormal=1)
| Contrast 135 < 0.353
| | Brightness > 0.448
| | Entropy > 5.362
| | | Contrast 45 > 0.388: Normal (Normal=2, Abnormal=0)
| | | Contrast 45 < 0.388
| | | | Entropy > 5.636
| | | | Brightness > 0.848
| | | | Entropy > 6.185: Abnormal (Normal=0, Abnormal=17)
| | | | Entropy < 6.185
| | | | | Contrast 45 > 0.284: Normal (Normal=3, Abnormal=1)
| | | | | Contrast 45 < 0.284
| | | | | | Contrast 45 > 0.249: Abnormal (Normal=1, Abnormal=2)
| | | | | | Contrast 45 < 0.249
| | | | | | | Contrast 135 > 0.245
| | | | | | | Entropy > 5.362: Abnormal (Normal=0, Abnormal=2)
| | | | | | | Entropy < 5.362: Normal (Normal=1, Abnormal=1)
| | | | | | | | Contrast 135 < 0.245: Abnormal (Normal=0, Abnormal=15)
| | | | | | | Brightness < 0.848: Abnormal (Normal=0, Abnormal=32)
| | | | | | | Entropy < 5.636
| | | | | | | | Brightness > 0.848: Normal (Normal=5, Abnormal=1)
| | | | | | | | Brightness < 0.848: Abnormal (Normal=0, Abnormal=3)
| | | | | | | | Entropy < 5.362: Normal (Normal=4, Abnormal=0)
| | | | | | | Brightness < 0.448
| | | | | | | | Contrast 135 > 0.185: Normal (Normal=15, Abnormal=0)
| | | | | | | | Contrast 135 < 0.185: Abnormal (Normal=0, Abnormal=5)
    
```

Gambar 12
Rule Hasil Klasifikasi Decesion Tree

Hasil dan Pengujian

Hasil dari proses klasifikasi dengan menggunakan decesion tree C45 akan diujikan dengan pengujian *Confusion Matrix*. Pengujian ini bertujuan untuk menguji berapa baik akurasi yang dihasilkan sistem dalam klasifikasi citra sel normal dan abnormal.

Tabel 2
Pengujian *Confusion Matrix*

		Kelas Hasil Prediksi	
		Normal	Abnormal
Kelas Aktual	Normal	59	1
	Abnormal	4	76

$$\text{Akurasi} = \frac{\text{Banyaknya Prediksi yang Benar}}{\text{Total Banyaknya Prediksi}} \times 100\%$$

Dari hasil perhitungan akurasi klasifikasi normal dan abnormal diatas, sistem yang dibuat mempunyai akurasi ketepatan dalam memprediksi sebesar 96,43% dan kesalahan dalam memprediksi 3,57%.

PENUTUP

Kesimpulan

Dalam penelitian ini diusulkan satu rangkaian proses analisis tekstur citra sel nukleus dan klasifikasi menggunakan algoritma C45 secara keseluruhan perbandingan hasil ekstraksi 18 tekstur dapat disimpulkan hal-hal sebagai berikut:

1. Tekstur dalam sel akan semakin jelas terlihat terutama pada saat pengujian kelas abnormal, tekstur sel abnormal lebih banyak terlihat dibandingkan dengan tekstur sel normal.
2. Dari hasil perhitungan akurasi klasifikasi citra nukleus normal dan abnormal akurasi yang diperoleh adalah 96,43% dan kesalahan dalam memprediksi adalah 3,57%.

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] Arifin T, Riana D, dan Hapsari, G.I. (2013) : Klasifikasi Statistikl Tekstur Sel Pap Smear Dengan Decesion Tree, Jurnal Informatika, Universitas BSI Bandung, **1**, 38.
- [2] Dalimartha S. (2004) : Deteksi Dini Kanker & Simplisia Antikanker. Jakarta: Penebar Swadaya Jakarta.
- [3] Gonzalez R.C, Woods R.E, & Eddins S.L. (2003) : *Digital Image Processing Using MATLAB*, 11-12
- [4] Haralick R.M, Shanmugan K, & Dinstein I. (2003) : Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, 610-621.
- [5] Indriyani C, & Riana D. (2010) : *Prediction Image Pap Smear Web*

- Based With Decision Tree*. STIMIK Nusa Mandiri , 1-5.
- [6] Jantzen J, Norup G.J, Dounias, & Bjerregaard B. (2005) : *Pap-smear Benchmark Data For Pattern Classification*, Technical University of Denmark, 1-20.
- [7] Muhimmah I, Anwariyah K, & Indrayanti. (2012) : *Extraction and Selection Features of Cervical Cell Types in Pap Smear Digital Images*. Wise Health ITB, 1-7.
- [8] Mathworks. (2012) : from Matrix Laboratory: <http://www.mathworks.com/>. (25 Desember 2012).
- [9] Martin E. (2003) : *Pap-Smear Classification. From Technical University of Denmark*: <http://labs.fme.aegean.gr/decision/downloads/> (25 Desember 2012).
- [10] Novitasari. (2010) : Analisis Identifikasi Serviks Normal dan Abnormal Berdasarkan Filter Gabor dan Ekstraksi Ciri Tekstur Statistik. Universitas Gunadarma , 1-7.
- [11] Prasetyo E. (2011) : Pengolahan Citra Digital Dan Aplikasinya Menggunakan Matlab. 1-2.
- [12] Pratama G, Riana D, & Hasanudin (2012) : *Pap Smear Nucleus Texture Analysis*. ITB , 1-4.
- [13] 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.
- [14] Selinger S. (2010) : *Image Procesing and Texture Analysis*. Dennis GaborCollege , 1-7. 1-20.
- [15] Suprpto. (2010) : Penggunaan Pengolahan Citra Digital Pada Pemeriksaan Pap Smear Dalam Pendeteksian Kanker Serviks. Universitas Brawijaya , 1-10.
- [16] WHO (2013) : WHO Guidance note. Number of pages 12 Publication 2013. From <http://www.who.int/reproductivehealth/publication/s/cancers/9789241505147/en/index.html> (19januari 2013).
- [17] Zuiderveld K. (2000) : Contrast Limited Adaptive Histogram Equalization. *Graphic Gems IV*. SanDiego: Academic Press Professional , 474-485.

Toni Arifin. Lahir di bandung 30 januari 1991, lulusan program sarjana (S1) Universitas BSI Bandung, saat ini aktif sebagai pengajar dan staf akademik di universitas BSI Bandung.