



## Identifikasi Citra Pap Smear RepoMedUNM dengan Menggunakan K-Means Clustering dan GLCM

Dwiza Riana<sup>1</sup>, Sri Rahayu<sup>2</sup>, Sri Hadiani<sup>3</sup>, Frieyadie<sup>4</sup>, Muhamad Hasan<sup>5</sup>, Izni Nur Karimah<sup>6</sup>, Rafly Pratama<sup>7</sup>

<sup>1</sup>Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

<sup>2,3,6,7</sup>Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

<sup>4,5</sup>Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

<sup>1</sup>dwiza@nusamandiri.ac.id, <sup>2</sup>sriahayu.rry@nusamandiri.ac.id\*, <sup>3</sup>sri.shv@nusamandiri.ac.id, <sup>4</sup>frieyadie@nusamandiri.ac.id,

<sup>5</sup>muhamad.mhx@nusamandiri.ac.id, <sup>6</sup>izninu12180531@nusamandiri.ac.id, <sup>7</sup>raflyp12180294@nusamandiri.ac.id

### Abstract

*Cervical cancer's a gynecological malignancy in women that's very dangerous, even causes death. Prevention through early detection of Pap smear test. It was carried out by pathologists with the help of a microscope still have obstacles in observations. There're many studies on Pap smear image processing for helping pathologists in cell identification. Availability of Pap smear image dataset is needed in cervical cancer early detection research. The purpose of this study was to segment, feature extraction and classify 180 Pap smear images of RepoMedUNM. The method used to identify Pap smear images begins with preprocessing, namely changing the color in the image to  $L^*a^*b$  color, segmentation using the K-means method, extraction of 6 features, namely metric, eccentricity, contrast, correlation, energy, and homogeneity, and then identified by calculating the closest distance between the training data features and the test data features with the Euclidean distance. The result of identification ThinPrep Pap smear images in 3 classes achieve average accuracy of 93.33%, Non-ThinPrep Pap smear images in 2 classes achieve 90% average accuracy and the average accuracy of the overall in the 4 classes reached 92%. These results indicate that the proposed method can identify Pap smear images well.*

*Keywords: Pap Smear, ThinPrep, Non-ThinPrep, RepoMedUNM, K-Means, GLCM*

### Abstrak

Kanker serviks merupakan keganasan ginekologi pada wanita yang sangat membahayakan bahkan menimbulkan kematian. Pencegahan melalui deteksi dini tes Pap smear. Pap smear dilakukan oleh ahli patologi dengan bantuan mikroskop dan masih memiliki kendala dalam pengamatannya. Banyak penelitian mengenai pengolahan citra Pap smear dengan tujuan membantu para ahli patologi dalam identifikasi sel. Ketersediaan dataset citra Pap smear sangat dibutuhkan pada penelitian deteksi dini kanker serviks. Tujuan penelitian ini adalah melakukan segmentasi, ekstraksi fitur dan klasifikasi pada 180 citra Pap smear RepoMedUNM. Metode yang digunakan untuk melakukan identifikasi citra Pap smear diawali dengan preprocessing, yaitu mengubah warna pada citra menjadi warna  $L^*a^*b$ , segmentasi menggunakan metode K-means, ekstraksi 6 fitur yaitu *metric, eccentricity, contrast, correlation, energy, dan homogeneity*, dan selanjutnya diidentifikasi dengan menghitung jarak terdekat antara fitur data latih dengan fitur data uji dengan persamaan euclidean distance. Hasil identifikasi citra Pap smear ThinPrep pada 3 kelas mencapai rata-rata akurasi 93,33%, sedangkan citra Pap smear Non-ThinPrep pada 2 kelas mencapai rata-rata akurasi 90% dan rata-rata akurasi keseluruhan citra pada 4 kelas mencapai 92%. Hasil ini menunjukkan bahwa metode usulan dapat mengidentifikasi citra Pap smear dengan baik.

Kata kunci: Pap Smear, ThinPrep, Non-ThinPrep, RepoMedUNM, K-Means, GLCM.

### 1. Pendahuluan

Sensus penduduk tahun 2020 mencatat jumlah penduduk berjenis kelamin wanita lebih sedikit, yaitu 133,54 juta orang atau 49,42 persen dari penduduk Indonesia[1]. Sebagian besar berada pada usia produktif. Kemenkes RI per Maret 2020 mencatat

kematian ibu-ibu di Indonesia mencapai 14.221 orang[2]. Kanker serviks merupakan penyakit kanker nomor dua terbanyak yang diderita wanita di dunia, yaitu sekitar 500.000 kasus baru dan kematian 250.000 setiap tahun[3]. Sedangkan di Negara berkembang, menjadi penyakit urutan pertama yang mematikan, hampir 80% kasus terjadi. Indonesia merupakan urutan

ke 8 di Asia Tenggara untuk kasus ini, terdapat 1362 jiwa wanita penderita kanker serviks[4].

Menurut *World Health Organization* (WHO) tahun 2018 terdapat 18,1 juta kasus kanker baru dan 9,6 juta kematian yang terjadi, diantaranya 23,4% kematian wanita di Indonesia diakibatkan oleh kanker serviks ini [3]. Bagaimana tidak, penyakit ini berasal dari tumbuhnya keganasan ginekologi atau tumor ganas pada leher rahim[5]. WHO menyebutkan bahwa hal ini dapat dicegah dengan cara diagnosi dini, skrining yang efektif dan program pengobatan[6].

Upaya pencegahan dilakukan dengan pemeriksaan sejak dini untuk kanker serviks dapat dilakukan dengan tes Pap smear[7]. Pemeriksaan Pap smear dapat diketahui seorang wanita mengidap penyakit kanker serviks atau tidak. Pap smear dilakukan oleh ahli patologi sesuai dengan keahliannya di bawah pengamatan mikroskop[8], namun masih terdapat kesulitan dalam pengamatan sel[9].

Kebanyakan dari peneliti terdahulu [10][11], menggunakan data publik yang sudah tersedia, sedangkan pada penelitian ini, agar penelitian berkembang dan *update*, *dataset* dibangun sendiri dengan melakukan akuisisi citra dan hasilnya disimpan pada *database* citra Pap smear yang diberi nama RepoMedUNM yang saat ini menyediakan 6.168 citra Pap smear [12][13].

*Machine learning* dalam perkembangan teknologi saat ini dapat menjadi solusi, salah satunya dengan teknik pengolahan citra[14]. Beberapa peneliti sebelumnya juga pernah melakukan teknik pengolahan citra Pap smear, diantaranya penelitian yang dilakukan dengan teknik segmentasi menggunakan metode *Markov Random Field* dan *K-means* pada citra kanker serviks[15]. Metode lain yaitu klasifikasi sel Pap smear menggunakan metode *Decision Tree* dengan optimasi *Particle Swarm Optimization* [16].

Metode K-means telah digunakan oleh peneliti sebelumnya yang bertujuan untuk mensegmentasi area *overlap* pada citra Pap smear dengan memanfaatkan fitur warna sel [17]. Persamaannya dengan penelitian ini adalah K-means juga digunakan dalam proses segmentasi tetapi bukan untuk daerah *overlapping* pada sel tetapi digunakan untuk mensegmentasi area nukleus dan sitoplasma.

Penggunaan metode K-means untuk segmentasi citra Pap smear dalam klasifikasi citra menghasilkan akurasi yang baik. Segmentasi K-means digunakan pada sel nukleus Pap smear yang menghasilkan 4 fitur *contrast*, *energy*, *entropy* dan *homogeneity* [18]. Penelitian tersebut menggunakan *dataset* Herlev sebanyak 28 citra sel tunggal yaitu kelas normal dan kelas abnormal. Selanjutnya klasifikasi citra dilakukan dengan menggunakan metode *Backpropagation Neural Network* (BNN) menghasilkan akurasi sebesar 88.8%

dan *error rate* sebesar 11.2%. Selain pada citra Pap smear, metode segmentasi K-means dengan ekstraksi ciri GLCM dan identifikasi dengan *euclidean distance* sudah digunakan untuk klasifikasi citra cacat kayu mahoni menghasilkan akurasi yang sangat baik yaitu 91,67% [19] dan akurasi 95,33% [20].

Penelitian pada 180 citra Pap smear RepomedUNM juga akan menggunakan segmentasi K-means untuk menghasilkan enam fitur tekstur yaitu *metric*, *eccentricity*, *contras*, *correlation*, *energy*, dan *homogeneity*. Klasifikasi dilakukan terhadap sebanyak 180 citra Pap smear RepomedUNM berkelompok yang tidak hanya terdiri dari kelas normal dan abnormal, tetapi terdiri dari Thinprep H-sil, Thinprep Koilocyt, Thinprep Normal, Non Thinprep L-sil, dan Non Thinprep Normal. Klasifikasi menggunakan *euclidean distance* pada citra ini dan menghasilkan akurasi yang baik.

## 2. Metode Penelitian

### 2.1. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini diambil dari dataset RepoMedUNM. Dataset ini menyediakan koleksi citra yang lebih bervariasi, yaitu citra Pap smear *ThinPrep* dan *Non-ThinPrep* (konvensional) dengan jumlah citra saat ini mencapai 6.168 citra sel Pap smear[12]. Citra dalam dataset RepoMedUNM ini diperoleh melalui hasil digitalisasi citra dengan menggunakan mikroskop optik OLYMPUS CX33RTFS2 dan mikroskop X52-107BN dan kamera Logitech.

Digitalisasi citra atau akuisisi citra adalah proses di mana kita memperoleh representasi digital dari sebuah adegan[21]. Hal-hal yang harus diperhatikan dalam proses akuisisi citra di antaranya adalah resolusi alat akuisisi, jarak dan sudut pengambilan citra, pencahayaan, perbesaran, pergerakan objek maupun pergerakan kamera, dan citra hasil akuisisi[22][23].



Gambar 1. Proses Akuisisi Citra

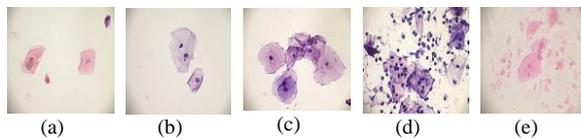
Gambar 1 menunjukkan proses digitalisasi citra Pap smear dengan mikroskop optik OLYMPUS CX33RTFS2 dan mikroskop X52-107BN dan kamera Logitech.

Tabel 1. Dataset yang Digunakan

No	Pap Smear	Kelas	Jumlah Citra
1	ThinPrep	H-sil	60
		Koilocyt	30
		Normal	30
2	Non-ThinPrep	L-sil	30
		Normal	30
Total			180

Pada tabel 1, dapat kita ketahui, *dataset* yang digunakan dalam penelitian ini sebanyak 180 citra yang terdiri dari Pap smear ThinPrep berjumlah 120 dengan 3 kelas, diantaranya H-sil, Koilocyt dan Normal, Pap smear Non-ThinPrep berjumlah 60 dengan 2 kelas, diantaranya L-sil dan Normal.

Gambar 2 merupakan contoh masing-masing kelas citra Pap smear sebagai berikut :



Gambar 2. Hasil Akuisisi Citra

Pada Gambar 2 diatas, dapat kita lihat contoh masing-masing citra yang digunakan dalam penelitian ini, yang mana Gambar 2.(a) adalah jenis Pap Smear ThinPrep kelas H-sil. Pada kelas H-sil, dysplasia parah, anisositosis, kromatin, bergranulasi atau nucleus lebih besar. Kelas H-sil merupakan Atipiae selular yang sering dikaitkan dengan infeksi persisten HPV dengan risiko tinggi berkembang menjadi kanker.

Gambar 2.(b) adalah jenis Pap smear ThinPrep Koilocyt. Sel koilocyt merupakan patognomonik dalam infeksi HPV yang biasanya menunjukkan derajat degenerasi yang bergantung pada berbagai virus yang berbeda. Koilocyt biasanya membesar dan terletak eksentrik, hiperkromatik serta menunjukkan ketidakteraturan pada membran nukelus. Pada beberapa kasus dalam koilocyt terdapat sel yang memiliki nucleus nya dua atau bahkan banyak.

Gambar 2.(c) adalah jenis Pap smear ThinPrep Normal, Sel Normal memiliki ciri relatif kecil, berbentuk bulat, dengan inti besar, terletak di tengah, sitoplasma sedikit, padat, agak gelap, dan basofilik. Sel ini juga adalah sel epitel skuamosa yang terletak di lapisan epitel.

Sedangkan, pada jenis Pap smear Non-ThinPrep, Gambar 2.(d) adalah L-Sil. L-Sil menunjukkan kondisi sel infeksi sementara dengan HPV. Sel ini berarti sel abnormal ringan berukuran sedang. Terakhir, Gambar 2.(e) jenis Pap smear Non-ThinPrep Normal, sel ini memiliki ciri ciri-ciri sel yang relatif kecil, bentuk bulat, dengan inti besar, terletak di tengah, sitoplasma sedikit, padat, sedikit gelap, dan basofilik Sel ini adalah sel epitel skuamosa yang terletak di lapisan epitel.

## 2.2. Identifikasi Citra Pap Smear

Judul Dalam penelitian ini, citra yang digunakan untuk diolah berjumlah 180 citra yang terdiri dari jenis ThinPrep dan Non-ThinPrep. Sebelum melakukan identifikasi, citra dikonversi warnanya terlebih dahulu, kemudian disegmentasi dan dikeluarkan nilai fiturnya dengan teknik ekstraksi fitur. Penelitian ini juga divisualisasikan dengan menggunakan aplikasi MatlabR2015a sehingga dapat mudah digunakan..

Dataset terlebih dahulu dibagi menjadi data latih dan data uji untuk melakukan pengujian apakah identifikasi dilakukan dengan tepat atau tidak. Data yang digunakan pada penelitian ini terdiri dari 120 citra Pap smear jenis ThinPrep dan 60 citra Pap smear jenis Non-ThinPrep. Masing – masing jenis dibagi ke dalam 60% data latih dan 40%-nya untuk data uji seperti yang terlihat pada Tabel 3 berikut :

Tabel 3. Data Citra Pap smear

No	Data Citra Pap smear		Jenis	Kelas	Jumlah Citra
	Data Citra Pap smear				
	Latih	Uji			
1	40	20	ThinPrep	H-sil	60
2	20	10		Koilocyt	30
3	20	10		Normal	30
4	20	10	Non-ThinPrep	L-sil	30
5	20	10		Normal	30
Total					180

Pada tabel 3 dapat kita lihat data citra Pap smear jenis ThinPrep yang digunakan berjumlah 120 citra dimana dari 3 kelas yaitu kelas H-sil, Koilocyt dan Normal ada 80 citra sebagai data latih yaitu kelas H-sil 40 citra, kelas Koilocyt dan Normal masing-masing 20 citra. Sedangkan data uji sebanyak 40 citra yaitu kelas H-sil 20 citra dan kelas Koilocyt dan Normal masing-masing 10 citra. Sedangkan data citra Pap smear jenis Non-ThinPrep yang digunakan berjumlah 60 citra dimana dari 2 kelas yaitu L-sil, dan Normal ada 40 citra sebagai data latih masing-masing kelas 20 citra dan data uji sebanyak 20 citra dengan masing-masing kelas 10 citra.

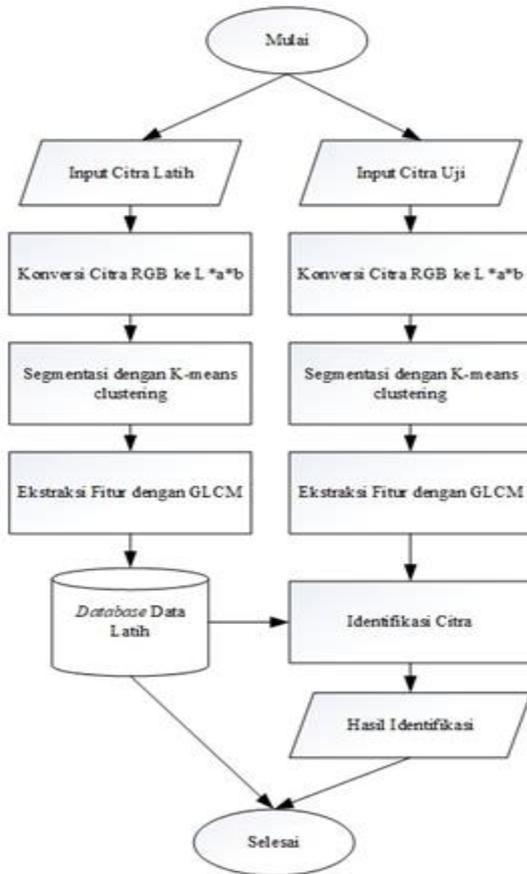
Terdapat beberapa teknik yang dilakukan peneliti untuk identifikasi citra Pap smear, yaitu *preprocessing* yang dilakukan dengan konversi warna dari RGB yaitu warna asli menjadi warna  $L^*a^*b$ . Proses segmentasi citra menggunakan metode K-means, ekstraksi fitur pada citra dengan 6 fitur GLCM yang terdiri dari *metric*, *eccentricity*, *contras*, *correlation*, *energy*, dan *homogeneity* serta identifikasi dengan mencari jarak paling pendek antara nilai-nilai fitur data uji terhadap nilai-nilai fitur data latih menggunakan *euclidean distance*.

## 2.3. Tahapan Penelitian

Tahapan-tahapan pada metode penelitian yang dilakukan disajikan pada Gambar 3.

### 1. Konversi Citra RGB ke L\*a\*b

Sebelum melakukan segmentasi, dengan metode k-means, citra data latih dikonversi ruang warnanya dari RGB ke L\*a\*b untuk mempersiapkan kegiatan segmentasi. Karena metode segmentasi k-means memiliki konsep mengelompokkan atau mengcluster dengan warna-warna terdekat, maka ruang warnanya harus lengkap dan ruang warna yang paling lengkap yang ditetapkan oleh Komisi Internasional tentang iluminasi warna (*French Commision Internationale de l'eclairage*) adalah ruang warna L\*a\*b[24][25].



Gambar 3. Metode Penelitian

### 2. Segmentasi dengan K-means Clustering

Segmentasi citra merupakan teknik mendasar dalam bidang dan pengolahan citra yang banyak diterapkan visi komputer. Dalam banyak kasus, ini dapat digambarkan sebagai tahap *preprocessing* dari pengenalan pola[26]. Segmentasi citra adalah proses membagi gambar menjadi kelas *pixel disjoint* berbeda, yang memiliki karakteristik yang mirip seperti tingkat abu-abu, warna, atau tekstur[27].

### 3. Ekstraksi Fitur dengan GLCM

Ekstraksi fitur yang dilakukan pada penelitian ini adalah ekstraksi fitur tekstur dan bentuk menggunakan

metode GLCM (*GrayLevel Co-occurrence Matrix*) seperti yang dilakukan [28] yang melakukan ekstraksi fitur setelah proses segmentasi untuk menganalisis tekstur dan bentuk serta mengeluarkan nilai pada 6 fitur diantaranya *metric, eccentricity, contras, correlation, energy, dan homogeneity* pada setiap citra dalam data latih sehingga terbentuklah *database* data latih.

### 4. Identikasi Citra

Langkah selanjutnya, setelah mendapatkan nilai ekstraksi fitur adalah melakukan identifikasi citra. Identifikasi ini dilakukan menggunakan persamaan *euclidean distance* yaitu menghitung jarak terdekat nilai fitur data uji terhadap nilai fitur data latih yang tersimpan di *database* data latih. Nilai terpendek merupakan nilai yang menunjukkan persamaan kelasnya.

*Euclidean Distance* merupakan teknik yang paling sederhana untuk menghitung jarak di antara dua vektor[29]. *Euclidean Distance* didefinisikan sebagai lurus jarak garis antara dua titik, yang memeriksa akar perbedaan kuadrat antara koordinat sepasang objek. Jarak *Euclidean* yang paling minimal menunjukkan jarak terdekat untuk proses pengenalannya[30].

Berikut ini rumus dari *Euclidean Distance* :

$$\begin{aligned}
 P &= (p_1, p_2, \dots, p_n) \\
 Q &= (q_1, q_2, \dots, q_n) \\
 d &= \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (1) \\
 &= \sqrt{\sum_{i=1}^n (p_i - q_i)^2}
 \end{aligned}$$

Dimana  $P$  adalah citra  $P$ ,  $Q$  adalah citra  $Q$ ,  $d$  adalah ukuran jarak antara *query* citra  $P$  dan  $Q$  yang ada di dalam database. Sedangkan  $p$  merupakan *feature vector* pada citra  $P$  dan  $q$  merupakan *feature vector* pada citra  $Q$ .

### 3. Hasil dan Pembahasan

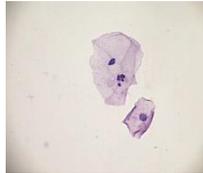
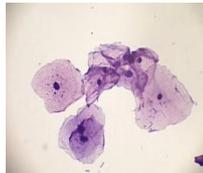
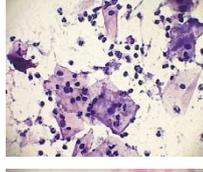
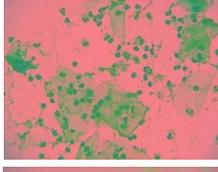
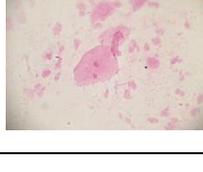
Berdasarkan tahapan-tahapan yang dilakukan pada metode penelitian, hasil yang dicapai dibahas dalam penjelasan berikut.

#### 3.1. Hasil Konversi Citra RGB ke L\*a\*b.

Konversi ruang warna RGB yang menjadi ruang warna citra asli ke dalam ruang warna L\*a\*b bertujuan untuk memberikan informasi warna lebih banyak terhadap citra. L berarti warna terang, \*a adalah singkatan dari koordinat merah/hijau, \*b adalah koordinat biru/kuning. Ruang warna ini memberi lebih banyak perbedaan antara warna satu sama lain karena perangkatnya yang independen[31]. Kelebihan dari model warna ini adalah tidak dipengaruhi oleh *device* pengambilan gambar. Hasil konversi warna RGB to L\*a\*b digambarkan pada Tabel 4.

Tabel 4 memberikan contoh hasil konversi warna asal yaitu RGB menjadi ruang warna  $L^*a^*b$ . Hal ini dilakukan untuk mempermudah segmentasi yang dilakukan dengan menggunakan *K-means*[17], sehingga *background*, sel sitoplasma dan nukleus mempunyai irama warna tersendiri, sehingga *K-means* akan dapat dengan mudah melakukan segmentasi pada tahap berikutnya.

Tabel 4. Hasil Konversi Warna RGB ke  $L^*a^*b$

Jenis/Kelas	Citra Asli	Citra $L^*a^*b$
<i>ThinPrep</i> / H-Sil		
<i>ThinPrep</i> / Koilocyt		
<i>ThinPrep</i> / Normal		
<i>Non-ThinPrep</i> / L-Sil		
<i>Non-ThinPrep</i> / Normal		

### 3.2. Hasil Segmentasi *K-means Clustering*

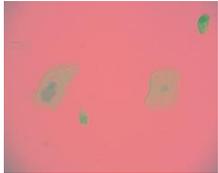
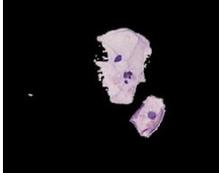
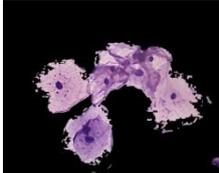
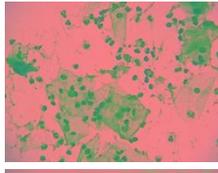
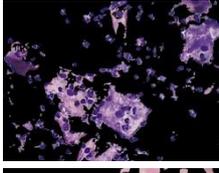
Segmentasi citra dengan *K-means* dilakukan dengan tujuan mendapatkan deskripsi dari sekumpulan data dengan cara mengungkapkan kecenderungan setiap data lainnya. Kecenderungan pengelompokan tersebut didasarkan pada kemiripan karakteristik data yang ada. Ide dasar dari teknik ini adalah menemukan pusat dari setiap kelompok data yang mungkin ada untuk kemudian mengelompokkan pada setiap data kedalam salah satu dari kelompok-kelompok tersebut berdasarkan jaraknya. Data dalam citra merupakan *pixel* dan intensitas warna yang ada didalamnya. *K-means* mengelompokkan berdasarkan kemiripan warna atau kedekatan nilai *pixel* sehingga mensegmentasi daerah yang sama warna atau teksturnya dari yang tidak sama. Hasil segmentasi citra menggunakan *k-means* digambarkan pada Tabel 5.

Pada Tabel 5, terlihat hasil segmentasi citra Pap *Smear Thinprep* dan *Non-Thinprep* dengan menggunakan *K-means* yang sebelumnya diubah ruang warnanya ke dalam ruang warna  $L^*a^*b$ . Segmentasi ini memisahkan bagian *background* dan *noise-noise* yang tidak penting dengan sel yang akan diolah. Pada segmentasi ini, yang diambil adalah sel yang akan diolah saja.

### 3.3. Hasil Ekstraksi Fitur

Setelah sel-sel yang akan diolah pada citra pap *smear* terpisah dari *background* dan *noise*, kemudian dilakukan ekstraksi fitur menggunakan GLCM (Gray Level Co-occurrence Matrix) pada 6 fitur metric, *eccentricity*, *contras*, *correlation*, *energy*, dan *homogeneity* sehingga terbentuklah database data latih.

Tabel 5. Hasil Segmentasi Citra dengan *K-means*

Jenis/Kelas	Citra $L^*a^*b$	Hasil Segmentasi
<i>ThinPrep</i> / H-Sil		
<i>ThinPrep</i> / Koilocyt		
<i>ThinPrep</i> / Normal		
<i>Non-ThinPrep</i> / L-Sil		
<i>Non-ThinPrep</i> / Normal		

Fitur-fitur ini digunakan karena dianggap cukup dalam melakukan identifikasi citra, seperti yang dijelaskan [32] bahwa *contras* menunjukkan ukuran elemen *spread* (momen inersia) dari matriks. Fitur *correlation* dapat memberikan petunjuk tentang struktur linear pada gambar. Fitur Energi mengekspresikan distribusi intensitas *pixel* ke tingkat abu-abu dan Homogenitas adalah nilai yang diperoleh dari kesamaan variasi dalam intensitas gambar. Sedangkan 2 fitur bentuk, Metrik adalah rasio antara area dan keliling dari area objek dan

*Eccentricity* adalah nilai rasio antara jarak fokus minor *elips* dengan fokus elips utama dari suatu area / bentuk pada objek.

Tabel 6. Nilai Rata-rata Hasil Ekstraksi Fitur Data Uji

	<i>Thin Prep/ H-Sil</i>	<i>Thin Prep/ Koilocyt</i>	<i>Thin Prep /Normal</i>	<i>NonThin inPrep /L-Sil</i>	<i>NonThin Prep / Normal</i>
<i>metric</i>	0.306	0.077	0.122	0.242	0.121
<i>eccentricity</i>	0.734	0.832	0.819	0.777	0.866
<i>contras</i>	0.433	1.162	0.760	0.804	0.669
<i>correlation</i>	0.659	0.321	0.318	0.318	0.315
<i>energy</i>	0.868	0.875	0.915	0.889	0.947
<i>homogeneity</i>	0.985	0.958	0.971	0.964	0.982

Dalam mengidentifikasi objek, hal penting yang harus diketahui adalah karakteristik yang dimiliki oleh objek. Tahapan ekstraksi fitur ini bertujuan untuk mengetahui karakteristik atau pola citra. Sehingga saat dikenali atau diidentifikasi, algoritma mencari karakteristik atau pola yang terdapat pada nilai-nilai fitur hasil ekstraksi. Pada Tabel 6 di atas, dapat diketahui nilai rata-rata hasil ekstraksi fitur seperti *metric*, *eccentricity*, *contras*, *correlation*, *energy*, dan *homogeneity* yang dihasilkan dari data uji.

### 3.3. Hasil Identifikasi Citra

Setelah melalui berbagai teknik pengolahan citra dan terbentuk *database* data latih, program dibuat menggunakan aplikasi MatlabR2015a. Data uji kemudian diidentifikasi dengan menghitung jarak terdekat antara nilai-nilai fitur pada data uji terhadap nilai-nilai fitur data latih pada *database* data latih dengan menghitung jarak terdekat terhadap keduanya menggunakan *euclidean distance*.

Data uji yang diidentifikasi juga melalui tahapan teknik-teknik yang sama seperti data latih yaitu konversi warna RGB ke  $L^*a^*b$ , segmentasi menggunakan *K-means* dan ekstraksi fitur menggunakan metode GLCM.

Tabel 7. Hasil Identifikasi

Jenis/Kelas	Hasil Identifikasi		Total Citra Uji
	Benar	Salah	
<i>ThinPrep/ H-Sil</i>	20	0	20
<i>ThinPrep/Koilocyt</i>	9	1	10
<i>ThinPrep/Normal</i>	9	1	10
<i>Non- ThinPrep /L-Sil</i>	9	1	10
<i>Non- ThinPrep /Normal</i>	9	1	10

Pada Tabel 7 disajikan hasil identifikasi data uji terhadap metode yang telah digunakan dengan teknik pengolahan citra sesuai bagan metode penelitian yang sudah dirancang sebelumnya.

Adapun untuk melihat seberapa jauh keberhasilan dari identifikasi ini, maka dibuatlah tabel *confussion matrix*. *Confussion matrix* berisi informasi tentang klasifikasi aktual dan prediksi yang dilakukan oleh *classifier*. Kinerja pengklasifikasi seperti itu biasanya dievaluasi menggunakan data dalam matriks[33][34]. Berikut

rumus perhitungan performasi dari *confussion matrix*[34]:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

Nilai TP (*True Positive*) diperoleh dari jumlah prediksi yang bernilai benar dan jumlah nilai asli yang benar untuk setiap *class* yang diuji. Nilai FN (*False Negative*) diperoleh dari jumlah prediksi yang bernilai salah dan jumlah nilai asli yang benar untuk setiap *class* yang diuji. Nilai FP (*False Positive*) diperoleh dari jumlah prediksi yang bernilai benar dan jumlah nilai asli yang salah untuk setiap *class* yang diuji. Nilai TN (*True Negative*) diperoleh dari jumlah prediksi yang bernilai salah dan jumlah nilai asli yang salah untuk setiap *class* yang diuji.

Tabel 8 merupakan tabel matriks *confussion matrix* beserta dengan prosentase keberhasilannya.

Tabel 8. *Confussion Matrix*

Jenis/ Kelas	Prediksi					Total
	<i>Thin Prep / H-Sil</i>	<i>Thin Prep/K oilocyt</i>	<i>Thin Prep /Nor mal</i>	<i>Non Thin Prep /L-Sil</i>	<i>Non Thin Prep /Nor mal</i>	
<i>ThinPrep / H-Sil</i>	20	0	0	0	0	20
<i>ThinPrep /Koilocyt</i>	0	9	1	0	0	10
<i>ThinPrep /Normal</i>	0	1	9	0	0	10
<i>Non- ThinPrep /L-Sil</i>	0	0	0	9	1	10
<i>Non- ThinPrep /Normal</i>	0	0	0	1	9	10

Pada Tabel 8, dapat kita ketahui *confussion matrix* yang terbentuk dari hasil pengujian program pada 3 kelas citra *Pap smear ThinPrep* dan 2 kelas pada citra *Pap smear Non-ThinPrep* yang digunakan dalam penelitian ini. Agar lebih mudah dalam melihat prosentase akurasi keberhasilannya.

Tabel 9. Akurasi

Kelas	Akurasi
<i>ThinPrep/ H-sil</i>	100%
<i>ThinPrep/Koilocyt</i>	90%
<i>ThinPrep/Normal</i>	90%
<b>Rata – Rata <i>ThinPrep</i></b>	<b>93,33%</b>
<i>Non- ThinPrep /L-sil</i>	90%
<i>Non- ThinPrep /Normal</i>	90%
<b>Rata – Rata <i>Non-ThinPrep</i></b>	<b>90%</b>
<b>Rata – Rata Total</b>	<b>92%</b>

Pada Tabel 9, dapat kita ketahui akurasi tingkat keberhasilan identifikasi pada masing-masing kelas, diantaranya *ThinPrep/H-sil* mendapatkan akurasi 100% dari total data uji 20 citra dan identifikasi benar 20 citra ( $20/20 = 1$ ), *ThinPrep/Koilocyt* mendapatkan akurasi

90% dari data uji 10 citra dan identifikasi benar 9 citra ( $9/10 = 0,9$ ), *ThinPrep/Normal* mendapatkan akurasi 90% dari data uji 10 citra dan identifikasi benar 9 citra ( $9/10 = 0,9$ ) sehingga dapat diambil rata-rata akurasi jenis citra Pap *smear ThinPrep* adalah sebesar 93,33%.

Pada citra Pap *smear Non-Thinprep*, kelas L-sil mendapatkan akurasi 90% dari data uji 10 citra dan identifikasi benar sebanyak 9 citra ( $0/10 = 0,9$ ) begitupun kelas normal pada jenis *Non-Thinprep* mendapat akurasi 90% dari data uji 10 citra dan identifikasi benar sebanyak 9 citra ( $0/10 = 0,9$ ). Sehingga dapat ditarik rata-rata akurasi citra Pap *smear* jenis *Non-Thinprep* adalah sebesar 90%. Setelah dilakukan identifikasi terhadap setiap kelas citra Pap *smear*, dapat diketahui kelas mana saja yang dapat diidentifikasi dengan baik, sehingga jika pada kelas yang kurang baik hasil identifikasinya dapat dilakukan penelitian lanjutan.

Hasil ini menunjukkan pula pada tiap-tiap kelas memiliki ciri fitur tekstur dan karakteristik unik. Sedangkan jika digabungkan, akurasi identifikasi citra Pap *smear* jenis *ThinPrep* dan *Non-ThinPrep* yang dilakukan pada 180 citra dari RepoMedUNM menggunakan segmentasi *k-means* dan ekstraksi fitur GLCM adalah sebesar 92%. Angka yang sangat baik untuk hasil identifikasi. Hasil ini menunjukkan bahwa secara keseluruhan 180 citra Pap *smear* dapat diklasifikasi dengan baik dengan metode yang digunakan, terlihat bahwa citra Pap *smear* baik *ThinPrep* maupun *Non-ThinPrep* memiliki nilai akurasi yang baik pada setiap kelas.

#### 4. Kesimpulan

Dari berbagai tahapan metode penelitian yang dilakukan peneliti menggunakan *dataset* RepoMedUNM. Identifikasi yang dilakukan pada 180 citra Pap *smear* untuk mengidentifikasi citra-citra tersebut secara tepat ke dalam tiap-tiap kelas. Identifikasi yang dilakukan menghasilkan rata-rata akurasi *ThinPrep* 93,33%, rata-rata akurasi *Non-ThinPrep* 90% dan rata-rata akurasi keseluruhan mencapai 92%. Hasil dari identifikasi ini sangat penting karena dapat menjadi pertimbangan klasifikasi citra Pap *smear* pada RepoMedUNM.

Semua tahapan penelitian yang telah dilakukan merupakan rangkaian dalam upaya membangun metode deteksi dini otomatis kanker serviks yang hingga saat ini masih merupakan penyakit ginekologi pada wanita yang sangat membahayakan bahkan menimbulkan kematian. Hasil klasifikasi pada citra Pap *smear* yang memiliki akurasi yang baik dapat mendukung upaya pembuatan sistem deteksi dini kanker serviks khususnya pada tes Pap *smear*. Sehingga dengan penelitian mengenai pengolahan citra Pap *smear* dapat membantu para ahli patologi dalam identifikasi sel.

Penelitian lanjutan dapat dilakukan dengan mengidentifikasi lebih lanjut citra Pap *smear* dengan berbagai fitur selain fitur tekstur. Selain itu, dapat dipertimbangkan metode klasifikasi lain seperti *deep learning* untuk mendapatkan hasil klasifikasi untuk keseluruhan citra baru dataset RepoMedUNM.

#### Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Direktorat Jenderal Pendidikan Tinggi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi yang telah mendukung penelitian ini melalui Hibah Penelitian Terapan Kompetitif Nasional “Repository Medical Imaging Citra Pap *Smear* untuk Deteksi Dini Cervical Cancer”, Tahun 2021.

#### Daftar Rujukan

- [1] A. N. Dzulfaroh, “Sensus Penduduk 2020: Jumlah Laki-Laki di Indonesia Lebih Banyak dari Perempuan,” *Kompas.com*, 2021. [Online]. Available: <https://www.kompas.com/tren/read/2021/01/22/113600465/sensus-penduduk-2020-jumlah-laki-laki-lebih-banyak-daripada-perempuan?page=all>.
- [2] O. Primadi, *Health Statistics (Health Information System)*, Katalog da. Jakarta: Kementrian Kesehatan RI, 2019.
- [3] J. H. Algadri, “Studi Deskriptif: Tindakan Pencegahan Kanker Serviks Pada Wanita Usia Subur di Wilayah Kerja Puskesmas Bakunase Kelurahan Fontein Kecamatan Kota Raja Kota Kupang,” *Compos. Part A Appl. Sci. Manuf.*, vol. 68, no. 1, pp. 1–12, 2020.
- [4] S. Haryani, D. Defrin, and Y. Yenita, “Prevalensi Kanker Serviks Berdasarkan Paritas di RSUP. Dr. M. Djamil Padang Periode Januari 2011- Desember 2012,” *J. Kesehat. Andalas*, vol. 5, no. 3, pp. 647–652, 2016.
- [5] S. Syaiful, F. L. Tarigan, and F. Zuska, “Skrining Kanker Serviks Dengan Pemeriksaan Pap Smear Pada Profesi Bidan Di Rumah Sakit Tk Ii Putri Hijau Medan Tahun 2017,” *J. Ris. Hesti Medan Akper Kesdam LBB Medan*, vol. 3, no. 2, p. 1, 2018.
- [6] N. A. Wantini and N. Indrayani, “Deteksi Dini Kanker Serviks dengan Inspeksi Visual Asam Asetat (IVA),” *J. Ners dan Kebidanan (Journal Ners Midwifery)*, vol. 6, no. 1, pp. 027–034, 2019.
- [7] A. N. Sholihah and E. Sulistyorini, “Hubungan Antara Sikap Pencegahan Kanker Serviks Dengan Minar Deteksi Dini Menggunakan Inspeksi Visual Asam Asetat Pada Wanita Usia Subur di RW IV Desa Cangkol Mojolaban Sukoharjo Tahun 2015,” *J. Kebidanan Indones.*, vol. 6, no. 2, pp. 102–116, 2018.
- [8] S. Anggraeni, “Self Efficacy Wanita Usia Subur untuk Melakukan Pap Smear Ditinjau dari Pengetahuan dan Dukungan Suami,” *Viva Med.*, vol. 10, no. 18, pp. 86–93, 2017.
- [9] R. Kurniawan, D. H. Fudholi, I. Muhimmah, A. Kurniawardhani, and Indrayanti, “Feature Analysis of Normal Epithelial Cervical Cell Characteristics in Pap Smear Images,” *2018 2nd Int. Conf. Informatics Comput. Sci. ICICoS 2018*, pp. 136–140, 2018.
- [10] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, “Pap-smear Benchmark Data For Pattern Classification,” *Proc. NiSIS 2005, Albufeira, Port.*, no. January 2006, pp. 1–9, 2005.
- [11] M. E. Plissiti *et al.*, “Sipakmed: A New Dataset For Feature And Image Based Classification Of Normal And Pathological Cervical Cells In Pap Smear Images Dept. of Computer Science & Engineering, University of Ioannina, Greece Dept. of Anatomy-Histology and Embryology, Facul.” *2018 25th IEEE Int. Conf. Image Process.*, pp. 3144–3148, 2018.
- [12] D. Riana *et al.*, “Repository Medical Imaging Citra Pap Smear,” <http://repomed.nusamandiri.ac.id/>, 2021. [Online].

- Available: <http://repomed.nusamandiri.ac.id/>.
- [13] D. Riana, S. Hadiani, and I. N. Karimah, "RepoMedUNM : A New Dataset for Feature Extraction and Training of Deep Learning Network for Classification of Pap Smear Images," *Int. Conf. Neural Inf. Process.*, vol. 1516, pp. 1–8, 2021.
- [14] T. Arifin, "Implementasi Algoritma PSO Dan Teknik Bagging Untuk Klasifikasi Sel Pap Smear," *J. Inform.*, vol. 4, no. 2, pp. 155–162, 2017.
- [15] R. S. D. Wijaya, Adiwijaya, Andriyan B Suksmono, and Tati LR Mengko, "Segmentasi Citra Kanker Serviks Menggunakan Markov Random Field dan Algoritma K-Means," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 139–147, 2021.
- [16] T. Arifin, "Optimasi Decision Tree menggunakan Particle Swarm Optimization untuk klasifikasi sel Pap Smear," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 3, pp. 572–579, 2020.
- [17] D. Riana, H. Tohir, and A. N. Hidayanto, "Segmentation of overlapping areas on pap smear images with color features using K-means and otsu methods," *Proc. 3rd Int. Conf. Informatics Comput. ICIC 2018*, pp. 1–5, 2018.
- [18] N. Putu, A. Oka, I. K. Gede, D. Putra, and K. S. Wibawa, "Klasifikasi Sel Nukleus Pap Smear Menggunakan Metode Backpropagation Neural Network," *J. Ilm. Merpat*, vol. 7, no. 3, pp. 224–232, 2019.
- [19] S. Rahayu, D. Riana, and Anton, "Segmentasi K-Means Dan Klasifikasi Cacat Pada Kayu Swietenia Mahagoni," *J. TECHNO Nusa Mandiri*, 2021.
- [20] D. Riana, S. Rahayu, M. Hasan, and Anton, "Comparison of segmentation and identification of swietenia mahagoni wood defects with augmentation images," *Heliyon*, vol. 7, no. 6, p. e07417, 2021.
- [21] G. P and V. Rajini, "YIQ Color Space based Satellite Image Segmentation using Modified FCM Clustering and Histogram Equalization," *Int. Conf. Adv. Electr. Eng.*, 2014.
- [22] V. K. Mishra, S. Kumar, and N. Shukla, "Image Acquisition and Techniques to Perform Image Acquisition," *SAMRIDDHI A J. Phys. Sci. Eng. Technol.*, vol. 9, no. 01, pp. 21–24, 2017.
- [23] S. Hadiani and D. Riana, "Segmentation and analysis of Pap smear microscopic images using the K-means and J48 algorithms," *J. Teknol. dan Sist. Komput.*, vol. 9, no. 2, pp. 113–119, 2021.
- [24] R. Rulaningtyas, A. B. Suksmono, T. Mengko, and P. Saptawati, "Multi Patch Approach in K-Means Clustering Method for Color Image Segmentation in Pulmonary Tuberculosis Identification," *Int. Conf. Instrumentation, Commun. Inf. Technol. Biomed. Eng.*, pp. 75–78, 2015.
- [25] K. Xiao *et al.*, "Characterising the variations in ethnic skin colours: a new calibrated data base for human skin," *Ski. Res. Technol.*, vol. 23, no. 1, pp. 21–29, 2017.
- [26] S. Hartanto, A. Sugiharto, and S. N. Endah, "Optical Character Recognition Menggunakan Algoritma Template Matching Correlation," *J. Informatics Technol.*, vol. 1, no. 1, pp. 11–20, 2012.
- [27] N. Dhanachandra, K. Mangle, and Y. J. Chanu, "Image Segmentation using K -means Clustering Algorithm and Subtractive Clustering Algorithm," *Procedia - Procedia Comput. Sci.*, vol. 54, pp. 764–771, 2015.
- [28] N. Merlina, E. Noersasongko, P. Nurtantio, M. A. Soeleman, D. Riana, and S. Hadiani, *Detecting the Width of Pap Smear Cytoplasm Image Based on GLCM Feature*. 2020.
- [29] F. U. Karimah, Ernawati, and D. Andreswari, "Rancang Bangun Aplikasi Pencarian Citra Batik Besurek Berbasis Tekstur Dengan Metode Gray Level Co- Occurrence Matrix Dan Euclidean DistanCE," *J. Teknol. Inf.*, vol. 11, no. April, pp. 64–77, 2015.
- [30] N. H. Barnouti and W. E. Matti, "Face Detection and Recognition Using Viola-Jones with PCA-LDA and Square Euclidean Distance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 5, pp. 371–377, 2016.
- [31] K. Raval, R. Shukla, and A. K. Shah, "Color Image Segmentation using FCM Clustering Technique in RGB , L \* a \* b , HSV , YIQ Color spaces," *Eur. J. Adv. Eng. Technol.*, vol. 4, no. 3, pp. 194–200, 2017.
- [32] I. Ihsan, E. W. Hidayat, and A. Rahmatulloh, "Identification of Bacterial Leaf Blight and Brown Spot Disease In Rice Plants With Image Processing Approach," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 5, no. 2, p. 59, 2020.
- [33] A. Bhardwaj and A. Tiwari, "Breast Cancer Diagnosis Using Genetically Optimized Neural Network Model," *Expert Syst. Appl.*, 2015.
- [34] A. Mardhotillah, A. Atfianto, and K. N. Rahmadhani, "Menghitung Jumlah Buah Cabe Berwarna Hijau Menggunakan Metode Transformasi Ruang Warna RGB," *e-Proceeding Eng.*, vol. 5, no. 2, pp. 3641–3648, 2018.