



Segmentasi Citra Kanker Serviks Menggunakan Markov Random Field dan Algoritma K-Means

Raihana Salsabila Darma Wijaya¹, Adiwijaya², Andriyan B Suksmono³, Tati LR Mengko⁴

^{1,2}Informatika, Fakultas Informatika, Universitas Telkom

^{3,4}Informatika, Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung

¹raihanawijaya@student.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id, ³suksmono@stei.itb.ac.id,

⁴tati@stei.itb.ac.id

Abstract

Cervical cancer is a dangerous disease caused by malignant tumors that grow on the cervix and has globally attacked many women. Pap smear test is one of the early prevention efforts for cervical cancer. Medical personnel often have difficulty identifying images of cervical cancer cells. Several studies have used the K-Means clustering method to identify cervical cancer cell images from Herlev dataset. This study uses the Herlev dataset with the K-Means clustering algorithm and also used the Markov Random Field parameter as a feature for the process of identifying cervical cancer cell images. This study compared the results of the proposed method with some differences in the preprocessing process. The experimental results show an accuracy of 74,51% for RGB channels without low pass filter. Accuracy of 75,11% is obtained from the segmentation process using RGB channels with low pass filter. A further increase in accuracy was obtained by 75,76% when the segmentation process used the grayscale channel with low pass Filter. Based on the segmentation experiment with the highest segmentation accuracy results, the classification process using K-Nearest Neighbor (KNN) gives an accuracy of 89,29%.

Keywords: cervical cancer, k-means clustering, k-nearest neighbor, markov random field, pap smear

Abstrak

Kanker serviks adalah penyakit berbahaya yang disebabkan oleh pertumbuhan tumor ganas pada serviks dan telah menyerang banyak wanita di dunia. Tes *pap smear* merupakan salah satu upaya pencegahan awal penyakit kanker serviks. Petugas medis seringkali kesulitan dalam mengidentifikasi citra sel kanker serviks. Beberapa penelitian telah menggunakan metode K-Means clustering untuk mengidentifikasi citra sel kanker serviks dari *dataset* Herlev. Penelitian ini juga menggunakan *dataset* Herlev dengan algoritma K-Means clustering, tetapi menggunakan parameter Markov Random Field sebagai fitur untuk proses pengidentifikasian citra sel kanker serviks. Penelitian ini melakukan perbandingan hasil dari metode yang diajukan dengan beberapa perbedaan pada *preprocessing*. Hasil eksperimen menunjukkan akurasi 74,51% untuk kanal RGB tanpa *low pass filter*, sedangkan akurasi sebesar 75,11% didapatkan dari proses segmentasi menggunakan kanal RGB dengan *low pass filter*. Peningkatan akurasi lebih lanjut didapatkan sebesar 75,76% saat proses segmentasi menggunakan kanal *grayscale* dengan *low pass filter*. Berdasarkan percobaan segmentasi citra dengan hasil akurasi segmentasi tertinggi, dilakukan proses klasifikasi menggunakan K-Nearest Neighbor (KNN) yang menghasilkan akurasi sebesar 89,29%.

Kata kunci: kanker serviks, k-means clustering, k-nearest neighbor, markov random field, pap smear

1. Pendahuluan

Kanker serviks merupakan jenis penyakit berbahaya yang menyerang banyak wanita di dunia. Penyakit ini terjadi pada wanita jika terdapat pertumbuhan tumor ganas di dalam serviks [1]. Lebih dari 99% kanker serviks mengandung *Human Papillomavirus* (HPV) yang dapat berasal dari: riwayat kehamilan; perilaku seksual; penggunaan kontrasepsi; merokok; nutrisi; dan

genetik [2]. Pada tahun 2018, diperkirakan 570.000 wanita didiagnosis kanker serviks di seluruh dunia dan sekitar 311.000 wanita meninggal karena penyakit tersebut [3]. Jumlah penderita kanker serviks di Indonesia sangat tinggi, setiap tahunnya tidak kurang dari 15.000 kasus [4]. Angka tersebut membuat kanker serviks merupakan penyakit pembunuh wanita nomor satu di Indonesia [1].

Diagnosa kanker serviks dapat ditegakkan melalui beberapa pemeriksaan laboratorium antara lain: 1) Pemeriksaan *pap smear* (sitologi), yaitu pemeriksaan dengan cara pengambilan lapisan dari permukaan leher rahim atau vagina untuk meninjau perubahan bentuk sel; 2) Pemeriksaan *Schiller* atau lebih dikenal dengan IVA, yaitu pemeriksaan dengan menggunakan larutan iodium untuk mengetahui perubahan warna jaringan yang mengalami kelainan; 3) Pemeriksaan kolposkopi, yaitu pemeriksaan dengan menggunakan alat untuk menentukan adanya daerah abnormal dan letak kelainannya [5].

Kanker serviks adalah salah satu bentuk kanker yang dapat disembuhkan jika terdeteksi dini. Kelainan jaringan pada serviks merupakan gejala awal kanker serviks, sehingga ketika mengalaminya dapat dilakukan upaya untuk mendeteksi dini, salah satunya dengan pemeriksaan *pap smear*. *Pap smear* adalah salah satu metode deteksi dini kanker serviks yang dinilai cukup akurat [5]. Pemeriksaan ini sebaiknya dilakukan setiap tahun, oleh wanita yang telah berusia 18 tahun atau telah aktif secara seksual walaupun tidak ada gejala kanker pada negara berkembang [1].

Pap smear adalah suatu metode pengambilan sel dari mulut rahim dengan alat khusus lalu dioleskan ke kaca objek yang diwarnai menggunakan metode *Papanicolaou* agar komponen sel dapat ditekankan oleh warna tertentu. *Slide* kaca yang dapat memuat hingga 300.000 sel dianalisis menggunakan mikroskop oleh sitoteknik secara manual [6]. Pekerjaan tersebut sangat memakan waktu dan dibutuhkan teknisi sitologi yang terampil dalam membedakan sel-sel yang beragam. Dengan adanya perkembangan sistem pengolahan citra medis diharapkan dapat mengurangi faktor subjektivitas dan kesalahan diagnosa [7].

Faktor keakuratan sistem pengolahan citra bergantung pada analisis citra *pap smear*. Analisis citra mengarah pada cara mengetahui tingkat keganasan sel abnormal yang dapat diwakili oleh fitur yang dihitung dari nukleus dan sitoplasma yang tersegmentasi, misalnya, rasio nukleus-sitoplasma atau fitur tekstur yang terkait dengan penyimpangan pola kromatin [8]. Namun, biasanya segmentasi nukleus lebih dapat diandalkan daripada sitoplasma (terdapat kemungkinan tumpang tindih pada bagian sitoplasma di citra multi-sel) [9].

Dalam 3 tahun terakhir ini, beberapa penelitian telah dilakukan terkait metode segmentasi citra sel kanker serviks dengan *dataset* dan penentuan kelas yang berbeda-beda. *Dataset* Herlev merupakan *dataset* yang umum digunakan dalam penelitian citra kanker serviks. Beberapa penelitian membagi jenis citra sel atas 2 kelas (normal dan abnormal) hingga 7 kelas (tingkatan sel abnormal). Berikut ini adalah beberapa penelitian sebelumnya terkait segmentasi citra sel kanker serviks yang ditunjukkan pada Tabel 1.

Tabel 1. Penelitian Terkait Mengenai Segmentasi Citra Sel Kanker Serviks

Penulis /Tahun	Data	Metode	Akurasi
W. Tanti et. al. / 2017 [10]	Herlev: 135 Citra	K-Means & Fuzzy C-Means, Klasifikasi Backpropagation (BNN)	2 Kelas: 84,44%
M. Sholik et. al. / 2017 [11]	Herlev: 7336 Citra (917×8)	Uniform Rotated Local Binary Pattern, Klasifikasi Fuzzy K-Nearest Neighbor	2 Kelas: 91,59% 7 Kelas: 67,89%
S. Gautam et. al. / 2018 [8]	Herlev: 917 Citra Aindra: 80 Citra	Patch-based CNN, Klasifikasi Decision-tree	2 Kelas: 99,3% 7 Kelas: 93,75%
Kurnianingsih et. al. / 2019 [12]	Herlev: 917 Citra	Regional Convolutional Neural Network (Mask R-CNN), Klasifikasi Visual Geometry Group-like Network (VGG-like Net)	2 Kelas: 98,1% 7 Kelas: 95,9%
N. Putu et. al. / 2019 [13]	Herlev: 28 Citra	K-Means & Morphological Operation, Klasifikasi Backpropagation (BNN)	2 Kelas: 88,8%

Berdasarkan Tabel 1 terdapat dua penelitian dengan penggunaan metode CNN pada proses segmentasi citranya, yaitu oleh S. Gautam et. al. dan Kurnianingsih et. al. Kedua penelitian ini menghasilkan akurasi yang tinggi yaitu sekitar 98-99%. Akurasi yang dihasilkan metode CNN akan semakin tinggi jika data uji yang digunakan berjumlah besar dan berlaku sebaliknya. Kedua penelitian ini menggunakan seluruh data citra pada *dataset* Herlev yang berjumlah 917 citra. Namun, jumlah data uji yang banyak tentu akan mempengaruhi waktu pemrosesan.

M. Sholik et. al. menggunakan metode *Uniform Rotated Local Binary Pattern* untuk proses segmentasi. Penelitian ini menggunakan seluruh data citra pada *dataset* Herlev yang setiap citranya dirotasi dalam delapan arah yaitu 0°, 45°, 90°, 135°, 180°, 225°, 270°, dan 315°. Sehingga jumlah citra yang digunakan menjadi 7336 citra. Nilai akurasi yang didapatkan cukup tinggi yaitu sebesar 91,59% menggunakan *Fuzzy KNN*[11].

W. Tanti et. al. dan N. Putu et. al. menggabungkan algoritma K-Means dan metode lain untuk proses segmentasi citra serta menggunakan metode BNN untuk proses klasifikasi. Kedua penelitian tersebut hanya menggunakan sebagian data citra pada *dataset* Herlev, yaitu 28 dan 135 citra. Untuk penggunaan data terbatas, penelitian tersebut menghasilkan akurasi yang baik sebesar 84,44% dan 88,8% [10].

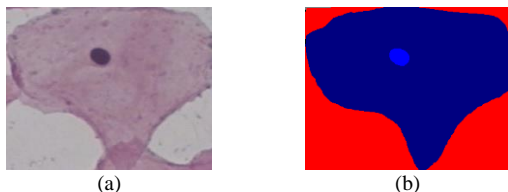
Tujuan dari penelitian ini adalah mengimplementasikan K-Means dengan parameter Markov Random Field (MRF) sebagai fitur dalam proses pengidentifikasian citra sel kanker serviks. Algoritma K-Means dalam proses segmentasi citra dipilih karena mampu menghasilkan akurasi yang cukup baik pada beberapa penelitian sebelumnya. Sedangkan parameter MRF dipilih untuk ekstraksi fitur karena dapat mengatasi jumlah data yang terbatas untuk proses pelatihan segmentasi citra menggunakan K-Means. Citra hasil segmentasi akan dibandingkan dengan citra acuan yang merupakan hasil dari segmentasi oleh dokter dari *dataset* Herlev untuk mendapatkan akurasi segmentasi. Selanjutnya, citra hasil segmentasi akan digunakan sebagai data uji dalam proses klasifikasi menggunakan algoritma KNN. Pemilihan algoritma KNN karena dapat mengatasi data yang terbatas dan dapat melatih banyak data *noise* dalam waktu yang relatif cepat [14].

2. Metode Penelitian

2.1. Data

Dataset Herlev berasal dari Rumah Sakit Universitas Herlev yang dapat diunduh dari <http://mde-lab.aegean.gr/index.php/downloads>. Teknisi sitologi dan dokter secara manual telah mengklasifikasi setiap citra sel ke dalam tujuh kelas. Setiap citra sel diklasifikasi oleh dua teknisi sitologi, sedangkan sampel citra sel yang sulit diklasifikasi oleh dokter. Jika terjadi ketidaksepakatan, sampel citra tidak digunakan agar *dataset* Herlev menyimpan diagnosis yang sepasti mungkin [6].

Terdapat tujuh kelas citra pada *dataset* Herlev, yaitu *normal superficial*, *normal intermediate*, *normal columnar*, *light dysplastic*, *moderate dysplastic*, *severe dysplastic*, dan *carcinoma in situ*. Data citra pada setiap kelas memiliki dua bentuk, yaitu citra *pap smear* dan citra hasil segmentasi dokter. Data latih yang digunakan sejumlah 10 citra dari setiap kelas, sehingga berjumlah 70 citra. Sedangkan untuk data ujinya digunakan sejumlah 4 citra dari setiap kelas, sehingga berjumlah 28 citra. Gambar 1 menunjukkan salah satu citra pada *dataset* Herlev.

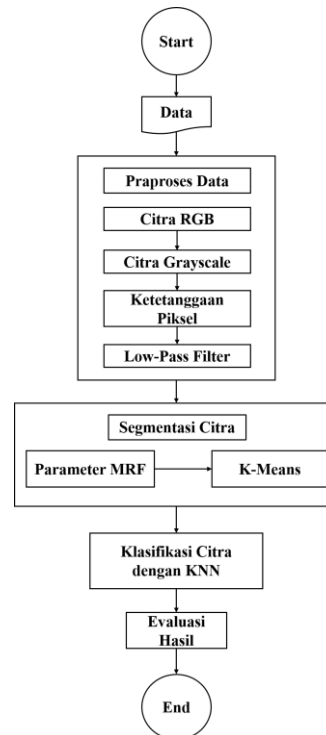


Gambar 1. Citra Pap Smear Kanker Serviks (a) Citra Kanker Serviks Hasil Segmentasi Dokter (b)

2.2. Skema Umum

Penelitian ini mengusulkan kombinasi fitur nilai keterkaitan setiap piksel pada citra dengan metode Markov Random Field (MRF) dan K-Means dalam segmentasi citra sel kanker serviks. Data citra yang

digunakan adalah *Dataset* Herlev. Skema umum penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Diagram Alir Penelitian

2.3. Praproses Data

Praproses citra dilakukan sebelum tahap segmentasi dan klasifikasi untuk mendapatkan fitur-fitur setiap jenis kelas citra serta evaluasi hasil. Praproses citra diawali dengan merubah citra RGB ke citra *grayscale*, lalu pada citra *grayscale* tersebut akan diterapkan *low pass filter* dengan sistem ketetapanan orde tiga.

2.3.1 Citra RGB

Setiap piksel pada citra berwarna mewakili warna yang merupakan kombinasi dari tiga warna dasar yaitu merah, hijau dan biru yang dikenal sebagai citra RGB (*Red*, *Green*, dan *Blue*). Setiap warna dasar mempunyai intensitas sendiri dengan nilai range 0-255. Sehingga, jumlah variasi warna yang dimiliki citra berwarna adalah $255^3 = 16.581.375$ (lebih dari 16 juta warna) [15].

2.3.2 Citra Grayscale

Citra aras keabuan memiliki satu nilai kanal disetiap piksel, dengan kata lain nilai kanal $red = green = blue$. Nilai tersebut menunjukkan tingkat intensitas. Tingkat intensitas dimulai dari warna hitam, keabuan sampai putih [16]. Proses konversi dari citra berwarna menjadi *grayscale* menggunakan koefisien dari ITU-Recommendation BT.709 [17].

$$Gray = 0.2125 \times R + 0.7154 \times G + 0.0721 \times B \quad (1)$$

2.3.3 Ketetangaan Pikel

Operasi ketetangaan piksel adalah operasi pengolahan citra untuk mendapatkan nilai suatu piksel yang melibatkan nilai piksel-piksel tetangganya. Hal ini didasarkan kenyataan bahwa setiap piksel pada umumnya tidak berdiri sendiri, melainkan terkait dengan piksel tetangga, karena merupakan bagian suatu objek tertentu di dalam citra. Sifat inilah yang kemudian mendasari timbulnya algoritma untuk mengolah setiap piksel citra melalui piksel-piksel tetangga [18]. Berikut ini Gambar 3 yang merupakan visualisasi dari ketetangaan piksel:

5	4	3	4	5
4	2	1	2	4
3	1	s	1	3
4	2	1	2	4
5	4	3	4	5

Gambar 3. Ketetangaan Pikel

2.3.4 Low-Pass Filter

Filtering atau penapisan merupakan suatu proses mengambil sebagian sinyal frekuensi tertentu dan membuang sinyal frekuensi yang lain. Bila ingin mempertahankan gradasi atau banyaknya *level* warna pada suatu citra, maka kita mempertahankan frekuensi rendah dan membuang frekuensi tinggi. Prinsip ini dinamakan *low pass filter* dan banyak digunakan untuk reduksi *blur* dan proses *blur*. *Low pass filter* adalah proses filter yang mengambil citra dengan gradasi intensitas halus dan perbedaan intensitas tinggi akan dikurangi atau dibuang. Pelembutan citra bertujuan untuk menekan gangguan (*noise*) pada citra [19]. Gambar 4 di bawah ini adalah visualisasi dari ketetangaan piksel orde tiga yang akan digunakan:

		t ₉		
	t ₁	t ₂	t ₃	
t ₁₂	t ₈	s	t ₄	t ₁₀
	t ₇	t ₆	t ₅	
		t ₁₁		

Gambar 4. Ketetangaan Pikel Orde Tiga

Nilai intensitas x_s didapatkan dari menjumlahkan nilai intensitas x_t dibagi dengan jumlah tetangganya yaitu n . Nilai intensitas x_s dihitung menggunakan persamaan berikut ini:

$$x_s = \frac{\sum_{i=1}^n x_{t_i}}{n} \quad (2)$$

2.4. Segmentasi Citra

Segmentasi citra adalah pemisahan objek yang satu dengan objek yang lain dalam sebuah citra. Dengan

proses segmentasi tersebut, masing-masing objek pada citra dapat diambil secara individu sehingga dapat digunakan sebagai input bagi proses lain [20]. Segmentasi memisahkan suatu citra ke dalam kelompok piksel-piksel homogen menjadi satu himpunan yang sesuai dengan masing-masing kriterianya. Tiap kelompok tersebut dinamakan segmen.

Setelah diperoleh *database* yang dihasilkan dari tahap praproses data, langkah selanjutnya adalah melatih data dengan menggunakan metode K-Means *clustering*. Langkah pertama untuk menggunakan metode K-Means *clustering* adalah dengan menentukan nilai K dengan menggunakan metode Elbow. Dalam penelitian ini, diperoleh nilai K sebesar 4. Setelah mendapatkan nilai K, dilakukan perhitungan jarak setiap data dengan *centroid*. Pada kasus ini, data dan *centroid* memiliki 14 dimensi sesuai dengan jumlah parameter yang dihasilkan dari tahap pengumpulan data. Kemudian hitung kembali nilai koordinat *centroid* tersebut hingga tidak berubah. Hasil yang diperoleh dari tahap pelatihan data ini adalah model yang berbentuk koordinat *centroid* sebanyak K titik.

Setelah dilakukan pelatihan dan mendapatkan model dari metode K-Means *clustering*, langkah selanjutnya adalah memprediksi data dengan menggunakan data pasien baru atau data uji. Langkah pertama pada tahap ini adalah dengan membaca data uji yaitu berupa citra kanker serviks. Lalu Citra tersebut dibaca dalam bentuk *grayscale*. Kemudian pada citra *grayscale* tersebut akan dihitung nilai rata-rata, variansi, dan perkalian piksel. Kemudian, nilai rata-rata, variansi, dan perkalian antar piksel yang bertetangga tersebut akan disimpan pada suatu variabel untuk diprediksi dengan model yang diperoleh dari tahap sebelumnya, yaitu pelatihan data. Hasil yang diperoleh dari tahap ini adalah citra nukleus dan citra sitoplasma, nilai 0 mewakili citra latar belakang dan nilai 1 mewakili citra nukleus serta citra sitoplasma tersebut. Akurasi segmentasi diperoleh dari pencocokan nilai piksel antara citra hasil segmentasi menggunakan algoritma dengan citra hasil segmentasi dari dokter.

2.4.1 Parameter Markov Random Field

Dalam suatu matriks 2-D jika setiap elemennya adalah sebuah variabel acak, maka matriks tersebut dinyatakan sebagai *discrete random fields* [21]. *Markov Random Field* (MRF) adalah model stokastik yang dapat diimplementasikan pada proses restorasi citra di berbagai bidang seperti multimedia, aplikasi medis, dan penginderaan jarak jauh [22]. Representasi suatu citra secara statistik akan mudah jika menggunakan pemodelan citra MRF [23]. Berdasarkan Gambar 4 posisi suatu piksel dinyatakan sebagai s dan nilai intensitas *grayscale* dinyatakan sebagai x_s . Luas suatu citra dinyatakan sebagai $M \times N$ dengan t adalah piksel tetangga dari s dan N_s adalah himpunan piksel tetangga

dari s sehingga nilai intensitas x_s bergantung pada piksel-piksel $t \in N_s$.

$$P(x_s | x_1, x_2, \dots, x_{s-1}, x_{s+1}, \dots, x_{M \times N}) = P(x_s | x_t) \quad (3)$$

Kondisi markov menyatakan bahwa probabilitas bersyarat dari piksel s memiliki nilai x_s hanya ditentukan dari piksel tetangganya, seperti yang dinyatakan pada persamaan (3). Klasifikasi dan segmentasi citra dapat dicapai dengan fitur parameter MRF [22]. Parameter yang digunakan pada penelitian ini berupa nilai rata-rata, variansi, dan perkalian intensitas s dengan $t \in N_s$. Nilai rata-rata didapatkan dari menjumlahkan nilai intensitas x_t dibagi dengan jumlah tetangganya yaitu n . Nilai rata-rata μ_s dihitung menggunakan persamaan berikut ini:

$$\mu_s = \frac{\sum_{i=1}^n x_{t_i}}{n} \quad (4)$$

Untuk nilai variansi σ^2 didapatkan dari menjumlahkan hasil selisih nilai intensitas x_t dan intensitas rata-rata μ_s yang dikuadratkan, kemudian dibagi dengan jumlah tetangganya. Nilai variansi yang menyatakan ukuran keragaman dari nilai piksel dapat ditentukan dengan persamaan berikut ini:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_{t_i} - \mu_s)^2}{n} \quad (5)$$

Sedangkan nilai parameter terakhir dinyatakan oleh vektor \vec{v} , yaitu perkalian dari intensitas piksel x_s dengan intensitas piksel $t \in N_s$ seperti pada persamaan berikut ini:

$$\vec{v} = (x_s \cdot x_{t_1}, x_s \cdot x_{t_2}, x_s \cdot x_{t_3}, \dots, x_s \cdot x_{t_{12}}) \quad (6)$$

2.4.2 K-Means

K-Means merupakan metode *unsupervised learning* yang digunakan untuk mengelompokkan suatu data menjadi beberapa kluster [2]. Metode K-Means akan mencari titik pusat kluster atau *centroid* melalui proses perulangan. Kedekatan atau kemiripan suatu data dengan data lain didasarkan pada seberapa dekat jarak data-data tersebut ke titik pusat kluster atau *centroid* [10]. Umumnya, pengelompokan data dengan K-Means dapat dilakukan dengan algoritma dasar sebagai berikut [24]:

1. Menentukan jumlah kluster k dan mengidentifikasi data $X_{ij} (i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, m)$ dengan n adalah jumlah piksel di dalam suatu citra yang akan dikluster dan m adalah jumlah parameter.
2. Menempatkan *centroid* sesuai dengan jumlah kluster C_k ditetapkan secara bebas untuk iterasi pertama.
3. Mengalokasikan data ke dalam *centroid* berdasarkan perhitungan jarak data X_{ij} dengan *centroid* C_k menggunakan formula Euclidean [17] seperti pada persamaan (7) berikut:

$$d = \sqrt{\sum_{i=1}^n (X_{ij} - C_{kj})^2} \quad (7)$$

Suatu data merupakan anggota dari kluster ke- k , jika jarak data ke pusat kluster ke- k bernilai paling kecil dibandingkan dengan jarak ke pusat kluster lainnya [17]. Hal ini dapat dihitung dengan menggunakan persamaan (8):

$$\min \sum_{k=1}^k d = \sqrt{\sum_{i=1}^n (X_{ij} - C_{kj})^2} \quad (8)$$

Menentukan titik *centroid* baru dengan menghitung rata-rata dari masing-masing data kluster saat ini [17]. Dimana X_{ij} adalah kluster ke- k dan p adalah banyaknya anggota kluster ke- k .

$$C_{kj} = \frac{\sum_{i=1}^p X_{ij}}{p} \quad (9)$$

Melakukan tahap 3 jika persamaan (9) masih menghasilkan nilai *centroid* yang berbeda dari nilai *centroid* sebelumnya, untuk kembali mengalokasikan data ke dalam kluster berdasarkan titik *centroid* baru.

Menghentikan iterasi jika persamaan (9) telah menghasilkan nilai *centroid* yang sama dengan nilai *centroid* sebelumnya.

2.5. Intersection over Union (IoU)

Intersection over Union (IoU) adalah nilai berdasarkan statistik keragaman sampel untuk mengevaluasi area tumpang tindih atau beririsan antara dua citra. Nilai IoU didapatkan dari perbandingan wilayah tumpang-tindih (*overlap area*) ($A \cap B$) dan wilayah gabungan (*union area*) ($A \cup B$) [25]. Jika A adalah citra segmentasi dokter dan B adalah citra hasil segmentasi dari suatu metode, maka akurasi segmentasi dapat dihitung dengan rumus IoU sebagai berikut:

$$IoU = \frac{A \cap B}{A \cup B} \quad (10)$$

2.6. Klasifikasi Citra dengan KNN

K-Nearest Neighbor (KNN) merupakan suatu metode yang menggunakan algoritma *supervised* [1]. KNN memberikan label kelas dari mayoritas pola terdekat-K dalam ruang data [2]. Pada tahap ini terdapat tiga fitur untuk melakukan klasifikasi, yaitu N (jumlah piksel citra nukleus), S (jumlah piksel citra sitoplasma), C (konstanta rasio antara nukleus dengan sitoplasma). Sedangkan untuk labelnya menggunakan variabel kelas citra (normal atau abnormal). Tahap-tahap untuk melakukan algoritma KNN pada penelitian ini adalah sebagai berikut:

1. Menyatakan parameter N , S , C , dan kelas. Parameter kelas berisi tujuh kelas pada tingkatan kanker serviks dan dinyatakan dalam bentuk angka. Diketahui terdapat citra berukuran $L \times W$ dengan nilai intensitas $I(x, y)$. Parameter N , S , dan C dihitung dengan persamaan berikut [6]:

$$N, S = \sum_{x=0}^L \sum_{y=0}^W I(x, y) \quad (11)$$

$$C = \frac{N}{(N+S)} \quad (12)$$

- Membuat *dataset* baru yaitu terdiri dari empat parameter (N , S , C , dan kelas) yang telah dinyatakan pada langkah 1,
- Menentukan besarnya K (jumlah tetangga),
- Menghitung jarak antara data uji dengan K jumlah data latih menggunakan rumus jarak Euclidian:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (13)$$

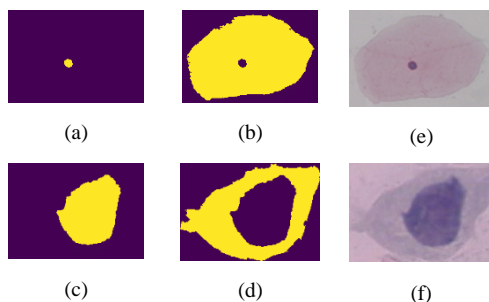
Variabel n pada persamaan (13) menyatakan jumlah parameter yang digunakan. Dalam kasus ini besarnya n yang digunakan adalah tiga (N , S , dan C). Variabel x_i dan y_i menyatakan koordinat x dan y dimensi ke- i .

- Mengurutkan jarak-jarak yang diperoleh dari hasil langkah 4 secara *ascending* (dari nilai rendah ke tinggi) dan mengambil K jumlah nilai terkecil.
- Klasifikasi akan selesai dengan mengambil jumlah kelas terbanyak dari langkah 5.
- Mengembalikan kategori dari kelas.
- Menghitung akurasi menggunakan rumus berikut:

$$a = \frac{\text{Jumlah Data Test Benar}}{\text{Jumlah Data Test Total}} \quad (14)$$

3. Hasil dan Pembahasan

Bagian ini akan menjelaskan tiga eksperimen dengan *preprocessing* yang berbeda dalam proses segmentasi, yaitu: (1) Kanal RGB tanpa *Low Pass Filter*, (2) Kanal RGB dengan *Low Pass Filter*, dan (3) *Grayscale* dengan *Low Pass Filter*. Tujuan dari perbandingan tersebut adalah untuk mengetahui kinerja dari tiap eksperimen. Gambar 5 di bawah merupakan salah satu citra sel kanker normal (*normal superficiel*) dan abnormal (*moderate dysplastic*) yang digunakan dalam penelitian ini.

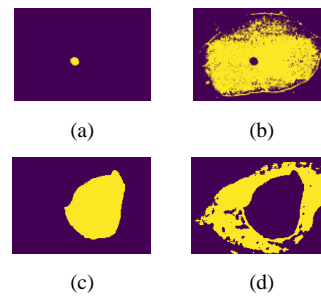


Gambar 5. Segmentasi Dokter Bagian Nucleus Normal (a) Segmentasi Dokter Bagian Sitoplasma Normal (b) Segmentasi Dokter Bagian Nucleus Abnormal (c) Segmentasi Dokter Bagian Sitoplasma Abnormal (d) Citra Pap Smear Kanker Serviks Kelas Normal (e) Citra Pap Smear Kanker Serviks Kelas Abnormal (f)

3.1. Segmentasi dengan Kanal RGB tanpa *Low Pass Filter*

Pada eksperimen ini Gambar 5e dan 5f akan dibaca dalam kanal RGB tanpa menggunakan *low pass filter*. Hasil dari *preprocessing* tersebut akan dilakukan ekstraksi fitur menggunakan parameter Markov Random Field. Jumlah fitur yang diperoleh dari proses ekstraksi

fitur sebanyak 42. Warna kuning pada Gambar 6 menunjukkan hasil segmentasi citra menggunakan kanal RGB tanpa *low pass filter* yang telah dipisahkan bagian nukleus dan sitoplasmanya.



Gambar 6. Bagian Nucleus Normal (a) Bagian Sitoplasma Normal (b) Bagian Nucleus Abnormal (c) Bagian Sitoplasma Abnormal (d)

Citra hasil segmentasi tersebut dibandingkan dengan citra yang sebelumnya telah disegmentasi oleh tenaga medis sehingga diperoleh hasil akurasi segmentasi citra sebesar 86,81% untuk bagian nukleus dan 60,12% pada citra sel kanker normal. Sedangkan pada citra sel kanker abnormal didapatkan akurasi sebesar 97,11% untuk bagian nukleus dan 71,70% untuk bagian sitoplasma.

Tabel 2. Hasil Akurasi Segmentasi dari Penggunaan Kanal RGB tanpa *Low Pass Filter*

Kelas	Nucleus	Sitoplasma
<i>Normal Superficiel</i>	74,34%	88,41%
<i>Normal intermediate</i>	65,68%	93,37%
<i>Normal Columnar</i>	77,10%	44,66%
<i>Light Dysplastic</i>	80,78%	68,93%
<i>Moderate Dysplatic</i>	75,14%	66,77%
<i>Severe Dysplastic</i>	70,52%	48,53%
<i>Carcinoma In Situ</i>	78,02%	39,48%
Rata-rata	74,51%	64,31%

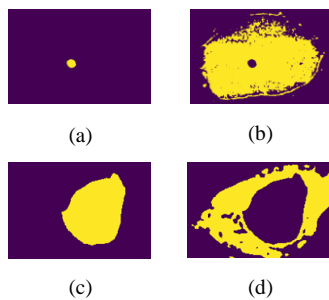
Hasil akurasi rata-rata segmentasi dari seluruh data uji ditunjukkan pada Tabel 2 di atas. Berdasarkan Gambar 7 citra hasil segmentasi dari nukleus memiliki banyak *noise* sehingga membuat akurasinya rendah. Akurasi dari penggunaan metode ini memiliki presentase rendah karena hanya menggunakan parameter RGB saja. Hal itu akan membuat proses segmentasi citra menggunakan K-Means *clustering* menjadi semakin detail, persebaran yang tidak merata, dan memiliki banyak *noise*.

3.2. Segmentasi dengan Kanal RGB dan *Low Pass Filter*

Pada eksperimen kedua, citra sel kanker serviks dibaca menjadi kanal RGB menggunakan *low pass filter*. Sama seperti eksperimen sebelumnya, dari proses ekstraksi fitur, didapatkan fitur sebanyak 42. Hasil segmentasi citra normal dan abnormal menggunakan kanal RGB dan *low pass filter* ditunjukkan pada Gambar 7 yang berwarna kuning.

Penggunaan *low pass filter* pada citra akan membuat persebaran nilai intensitas dari citra tersebut menjadi lebih merata. Dapat dilihat juga dari Gambar 7 bahwa citra hasil segmentasi menggunakan K-Means *clustering*

dari nukleus memiliki *noise* yang lebih sedikit jika dibandingkan dengan tidak digunakannya *low pass filter*. Jika dibandingkan dengan citra yang sebelumnya telah disegmentasi oleh tenaga medis, didapatkan hasil akurasi segmentasi citra sebesar 86,42% untuk bagian nukleus dan 58,53% pada citra sel kanker normal. Sedangkan pada citra sel kanker abnormal didapatkan akurasi sebesar 96,39% untuk bagian nukleus dan 72,48% untuk bagian sitoplasma. Hasil akurasi rata-rata segmentasi dari seluruh data uji ditunjukkan pada Tabel 3.



Gambar 7. Bagian Nukleus Normal (a) Bagian Sitoplasma Normal (b) Bagian Nukleus Abnormal (c) Bagian Sitoplasma Abnormal (d)

Tabel 3. Hasil Akurasi Segmentasi dari Penggunaan Kanal RGB dan *Low Pass Filter*

Kelas	Nukleus	Sitoplasma
<i>Normal Superficial</i>	77,81%	94,13%
<i>Normal intermediate</i>	70,89%	92,34%
<i>Normal Columnar</i>	79,42%	53,91%
<i>Light Dysplastic</i>	85,53%	69,69%
<i>Moderate Dysplastic</i>	85,53%	65,43%
<i>Severe Dysplastic</i>	73,25%	49,84%
<i>Carcinoma In Situ</i>	78,69%	36,55%
Rata-rata	75,11%	61,32%

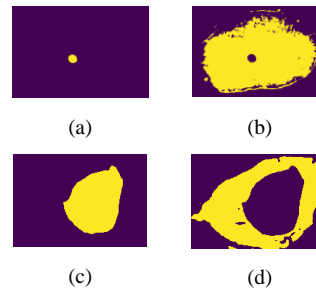
Pada eksperimen ini, didapatkan peningkatan akurasi rata-rata pada beberapa kelas jika dibandingkan dengan tidak digunakannya *low pass filter*. Peningkatan akurasi rata-rata terdapat pada segmentasi citra bagian nukleus dan pada segmentasi citra bagian sitoplasma terdapat penurunan akurasi rata-rata.

3.3. Segmentasi dengan *Grayscale* dan *Low Pass Filter*

Pada eksperimen ketiga, citra sel kanker serviks dibaca menjadi kanal *grayscale* dan menggunakan *low pass filter*. Jumlah fitur yang diperoleh dari proses ekstraksi fitur menggunakan Markov Random Field sebanyak 14. Jumlah fitur pada eksperimen ini memiliki nilai yang lebih kecil jika dibandingkan dengan eksperimen sebelumnya sehingga proses segmentasi berlangsung lebih cepat. Hasil segmentasi citra normal dan abnormal menggunakan kanal *grayscale* ditunjukkan pada Gambar 8 yang berwarna kuning.

Pada eksperimen ini, didapatkan hasil citra segmentasi yang lebih sedikit *noise*-nya jika dibandingkan dengan eksperimen sebelumnya. Penggunaan kanal *grayscale* menghasilkan hasil segmentasi citra tidak terlalu detail. Proses segmentasi citra menggunakan kanal *grayscale*

juga menghasilkan waktu pemrosesan lebih cepat karena memiliki jumlah parameter yang lebih sedikit jika dibandingkan dengan jumlah parameter pada kanal RGB. Didapatkan akurasi segmentasi citra sebesar 89,03% untuk bagian nukleus dan 76,77% pada citra sel kanker normal. Sedangkan pada citra sel kanker abnormal didapatkan akurasi sebesar 97,40% untuk bagian nukleus dan 77,38% untuk bagian sitoplasma. Untuk eksperimen ketiga ini, didapatkan hasil akurasi yang lebih baik dari eksperimen-eksperimen sebelumnya. Hasil akurasi rata-rata segmentasi untuk seluruh data uji ditunjukkan pada Tabel 4.



Gambar 8. Bagian Nukleus Normal (a) Bagian Sitoplasma Normal (b) Bagian Nukleus Abnormal (c) Bagian Sitoplasma Abnormal (d)

Tabel 4. Hasil Akurasi Segmentasi dari Penggunaan Kanal *Grayscale* dan *Low Pass Filter*

Kelas	Nukleus	Sitoplasma
<i>Normal Superficial</i>	52,24%	86,83%
<i>Normal intermediate</i>	74,57%	92,69%
<i>Normal Columnar</i>	79,55%	48,22%
<i>Light Dysplastic</i>	85,61%	66,46%
<i>Moderate Dysplastic</i>	70,14%	70,77%
<i>Severe Dysplastic</i>	81,42%	57,23%
<i>Carcinoma In Situ</i>	86,83%	42,79%
Rata-rata	75,76%	66,43%

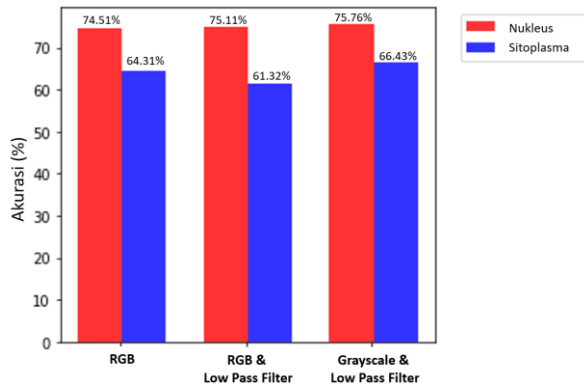
Sama seperti eksperimen sebelumnya, terdapat peningkatan akurasi pada beberapa kelas. Namun, eksperimen segmentasi menggunakan kanal *grayscale* dan *low pass filter* mendapatkan akurasi rata-ratanya tertinggi jika dibandingkan dengan akurasi rata-rata pada eksperimen-eksperimen sebelumnya.

3.4. Analisis Klasifikasi dengan KNN

Klasifikasi menggunakan KNN dilakukan pada hasil segmentasi citra dari eksperimen terbaik yaitu dengan penggunaan kanal *grayscale* dan *low pass filter*. Akurasi terbaik didapatkan sebesar 75,76% untuk bagian nukleus dan sebesar 66,43% untuk bagian sitoplasma. Gambar 9 menunjukkan penggunaan kanal *grayscale* dan *low pass filter* menghasilkan rata-rata akurasi segmentasi tertinggi dari eksperimen-eksperimen yang telah dilakukan.

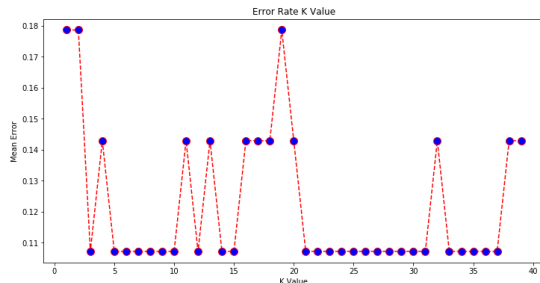
Untuk proses klasifikasi digunakan 70 citra segmentasi dokter pada *dataset* Herlev sebagai data latih dan 28 citra hasil segmentasi pada proses sebelumnya sebagai data uji. Masing-masing data citra dihitung nilai *N*

(jumlah piksel citra nukleus), S (jumlah piksel citra sitoplasma), C (konstanta rasio antara nukleus dengan sitoplasma) sebagai tiga fitur yang digunakan dalam proses klasifikasi. Sedangkan untuk label yang digunakan adalah variabel kelas citra dengan dua nilai yaitu normal (*normal superficial, normal intermediate, normal columnar*) atau abnormal (*light dysplastic, moderate dysplastic, severe dysplastic, carcinoma in situ*).



Gambar 9. Rata-rata Hasil Akurasi Segmentasi dari Eksperimen 1-3

Pada proses klasifikasi menggunakan algoritma KNN, jumlah tetangga atau nilai K yang digunakan sebesar tiga. Dipilihnya jumlah tetangga tersebut karena memiliki tingkat kesalahan (*mean error*) yang paling kecil dan paling optimal. Gambar 10 berikut ini menunjukkan pengaruh nilai K terhadap sebaran *mean error*.



Gambar 10. Pengaruh Nilai K terhadap Sebaran Mean Error

Akurasi klasifikasi didapatkan dengan menghitung jumlah data uji yang benar dibagi dengan total keseluruhan data uji. Klasifikasi menggunakan KNN menghasilkan nilai akurasi sebesar 89,29%. Penelitian sebelumnya oleh W. Tanti et. al. dan N. Putu et. al. juga menggunakan *dataset* Herlev dan menggunakan metode K-Means untuk proses segmentasi citranya. Hasil akurasi klasifikasi 2 kelas dari penelitian sebelumnya terdapat pada Tabel 5.

Berdasarkan Tabel 5 proses klasifikasi menggunakan KNN menghasilkan akurasi yang baik. Pada kasus ini, algoritma KNN dapat mengatasi jumlah data yang terbatas dan dapat melatih data yang memiliki *noise* dalam waktu yang relatif cepat [14]. Namun penggunaan algoritma KNN perlu memperhatikan besarnya nilai K ,

tetapi hal tersebut dapat diatasi dengan melihat sebaran *mean error* untuk mengetahui nilai K yang terbaik.

Tabel 5. Hasil Akurasi Klasifikasi 2 Kelas

Metode	Data Latih	Data Uji	Akurasi
K-Means & Fuzzy C-Means, Klasifikasi BNN [10]	90	45	84,44%
K-Means & Morphological Operation, Klasifikasi BNN [13]	19	9	88,88%
K-Means & MRF, Klasifikasi KNN	70	28	89,29%

4. Kesimpulan

Penelitian ini telah memberikan hasil akurasi yang baik dalam proses segmentasi citra sel kanker serviks menggunakan K-Means dan MRF serta proses klasifikasi untuk kasus data terbatas menggunakan KNN. Akurasi hasil segmentasi citra dengan preprocessing menggunakan kanal *grayscale* dan *low pass filter* memperoleh hasil tertinggi dibandingkan eksperimen preprocessing lainnya, yaitu sebesar 75,76% untuk nukleus dan 66,43% untuk sitoplasma. Sedangkan akurasi hasil klasifikasi dua kelas menggunakan KNN diperoleh hasil sebesar 89,29%.

Daftar Rujukan

- [1] 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," *Jurnal Riset Hesti Medan*, vol. 3, no. 2, pp. 1–15, 2018.
- [2] S. Rio and E. S. T. Suci, "Persepsi tentang Kanker Serviks dan Upaya Prevensinya pada Perempuan yang Memiliki Keluarga dengan Riwayat Kanker," *Jurnal Kesehatan Reproduksi*, vol. 4, no. 3, pp. 159–169, 2017.
- [3] A.n., "Cervical Cancer," *WHO*, 2018. [Online]. Available: https://www.who.int/health-topics/cervical-cancer#tab=tab_1. [Accessed: 08-Oct-2020].
- [4] H. Latifah, E. Nurachmah, and Hiryadi, "Menjalani Pemeriksaan PAP Smear Pasien Kanker Serviks di Poli Kandungan," *Jurnal Keperawatan Suaka Insan*, vol. 5, no. 1, pp. 90–99, 2020.
- [5] Y. Kusumawati, R. W. Nugrahaningtyas, and E. N. Rahmawati, "Pengetahuan, Deteksi Dini dan Vaksinasi HPV sebagai Faktor Pencegah Kanker Serviks di Kabupaten Sukoharjo," *Jurnal Kesehatan Masyarakat*, vol. 11, no. 2, pp. 204–213, 2016.
- [6] E. Martin, "Pap-Smear Classification," p. 101, 2003.
- [7] Rahmadwati, "Sistem Diagnosis Kanker Servik Berdasarkan Karakteristik Morfologi," *Jurnal EECIS*, vol. 7, no. 2, pp. 191–6, 2014.
- [8] S. Gautam, H. K. K., N. Jith, A. K. Sao, A. Bhavsar, and A. Natarajan, "Considerations for a PAP Smear Image Analysis System with CNN Features," pp. 1–8, 2018.
- [9] N. P. Husain and C. Fatichah, "Segmentasi Citra Sel Tunggal Smear Serviks Menggunakan Radiating Component Normalized Generalized GVFS," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 6, no. 1, pp. 107–114, 2017.
- [10] W. D. Tanti, E. Purwanti, and A. Supardi, "Identifikasi Kanker Serviks Dari Citra Papsmear Berbasis Kecerdasan Buatan Winda," *Jurnal Fisika dan Terapannya*, vol. 3, no. 3, pp. 98–111, 2015.
- [11] M. Sholik and C. Fatichah, "Klasifikasi Sel Serviks Pada Citra Pap Smear berdasarkan Fitur Bentuk Deskriptor Regional dan

- Fitur Tekstur Uniform Rotated Local Binary Pattern,” *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 15, no. 2, p. 214, 2017.
- [12] Kurnianingsih *et al.*, “Segmentation and classification of cervical cells using deep learning,” *IEEE Access*, vol. 7, no. August, pp. 116925–116941, 2019.
- [13] N. Putu, A. Oka, I. K. Gede, D. Putra, and K. S. Wibawa, “Klasifikasi Sel Nukleus Pap Smear Menggunakan Metode Backpropagation Neural Network,” *Jurnal Ilmiah Merpati*, vol. 7, no. 3, pp. 224–232, 2019.
- [14] S. D. Jadhav and H. P. Channe, “Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques,” *International Journal of Science and Research (IJSR)*, vol. 5, no. 1, pp. 1842–1845, 2016.
- [15] S. H. Wibowo and F. Susanto, “Penerapan Metode Gaussian Smoothing Untuk,” *Jurnal Media Infotama*, vol. 12, no. 2, pp. 129–135, 2016.
- [16] F. Riandari, “Implementasi Metode Geometric Mean Filter Untuk Perbaikan Dengan Reduksi Noise Pada Citra Digital,” *Jurnal Mantik Penusa*, vol. 2, no. 2, pp. 175–179, 2018.
- [17] I. A. Kesuma, Herman, and Munawir, “Penerapan Metode Kluster K-Means Pada Segmentasi Warna Citra,” *Seminar Nasional Inovasi dan Teknologi Informasi*, vol. 1, no. 3, pp. 427–430, 2016.
- [18] H. Pangaribuan, “Optimalisasi Kualitas Citra Digital Dengan Metode Ketetapan Piksel,” *Jurnal Ilmiah Informatika*, vol. 7, no. 01, p. 18, 2019.
- [19] M. Nugraheni, “Aplikasi Transformasi Watershed Untuk Segmentasi Citra Dengan Spatial Filter Sebagai Pemroses Awal,” *Seminar Nasional Informatika 2010*, vol. 1, no. 1, pp. 76–81, 2010.
- [20] P. A. Cahyan, M. Aswin, and A. Mustofa, “Segmentasi Citra Digital dengan Menggunakan Algoritma Watershed dan Lowpass Filter sebagai Proses Awal,” *Jurnal Mahasiswa TEUB*, vol. 1, no. 1, pp. 403–494, 2013.
- [21] Y. Antawiryawan, S. Violina, and A. Romadhony, “Analisis Segmentasi Citra Tuberculosis Menggunakan Markov Random Field,” pp. 1–6, 2011.
- [22] A. B. Suksmono and A. Hirose, “Adaptive noise reduction of InSAR images based on a complex-valued MRF model and its application to phase unwrapping problem,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 699–709, 2002.
- [23] A. Kusworo and A. B. Suksmono, “Pembangkitan Dan Pemulihan Citra Biner Markov Random Field (MRF) Secara Stokastik Dengan Algoritma Markov Chain Monte Carlo (MCMC),” *Berkala Fisika*, vol. 12, no. 4, pp. 145–152, 2009.
- [24] F. G. Febrinanto, C. Dewi, and A. T. Wiratno, “Implementasi Algoritme K-Means Sebagai Metode Segmentasi Citra Dalam Identifikasi Penyakit Daun Jeruk,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, vol. 2, no. 11, pp. 5375–5383, 2018.
- [25] H. Rezatofighi, N. Tsoi, J. Y. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” *arXiv*, pp. 658–666, 2019.