

SEGMENTASI DAN PEMISAHAN SEL DARAH PUTIH BERSENTUHAN MENGGUNAKAN K-MEANS DAN HIERARCHICAL CLUSTERING ANALYSIS PADA CITRA LEUKEMIA MYELOID AKUT

Aryo Harto¹⁾, Chastine Fatichah²⁾

^{1, 2)} Teknik Informatika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
e-mail: aryo14@mhs.if.its.ac.id¹⁾, chastine.fatichah@gmail.com²⁾

ABSTRAK

Keberhasilan identifikasi dan klasifikasi pada proses diagnosis penyakit Acute Myeloid Leukemia (AML) berbasis pengolahan citra sangat bergantung pada hasil segmentasi. Segmentasi pada citra mikroskopis sediaan apus darah tepi bertujuan untuk memisahkan region sel darah putih dari region lainnya. Untuk meningkatkan akurasi segmentasi pada citra AML beberapa hal seperti efek pencahayaan, kontras, staining yang beragam dan keberadaan sel yang bersentuhan harus diatasi. Pada penelitian ini diusulkan metode segmentasi dan pemisahan sel darah putih bersentuhan pada citra AML menggunakan analisis cluster dengan metode K-Means dan hierarchical clustering analysis (HCA). Segmentasi dilakukan menggunakan analisis cluster dengan metode K-Means pada citra AML yang memiliki variasi staining beragam. Keberadaan sel yang bersentuhan dipisahkan dengan metode HCA secara otomatis dengan hasil cluster yang stabil. Segmentasi dan pemisahan sel akan dilakukan pada masing-masing region lokal atau sub-citra yang didapatkan dari cropping pada citra AML. Untuk mendeteksi sel yang bersentuhan digunakan fitur geometri berupa kebulatan dan luasan. Dari hasil evaluasi yang dilakukan terhadap 40 citra pada dataset AML, metode yang diusulkan melakukan segmentasi sel darah putih dengan baik dan memisahkan touching cell menjadi sel-sel tunggal. Nilai rata-rata dari hasil segmentasi masing-masing adalah 0,975 untuk presisi, 0,885 untuk recall dan 0,928 untuk Zijdenbos similarity index (ZSI) pada region sel darah putih. Sedangkan pada region nukleus nilai rata-rata presisi adalah 0,975, recall adalah 0,924 dan ZSI adalah 0,948. Pada perhitungan jumlah sel, kesalahan yang dihasilkan juga cukup rendah yaitu sebesar 7,68%.

Kata Kunci: AML, clustering, hierarchical clustering analysis, k-means, segmentasi

ABSTRACT

The success of identification and classification on diagnosing acute myeloid leukemia (AML) diseases based on image processing relies heavily on segmentation result. Segmentation on peripheral blood smear images aims to separate the leukocytes region with others region. To increase the segmentation accuracy on AML images, a few things regarding lighting condition, contrast, staining variations and the existence of touching cells must be overcome. In this study a method for leukocytes segmentation and separate the touching cell on AML images using cluster analysis with K-Means and hierarchical clustering analysis (HCA) is proposed. K-Means method is used to analyze the cluster for AML images segmentation. The AML image datasets with various staining variations is segmented using K-Means method. The existence of touching cells is separated using HCA method which produce a stable clusters result. Segmentation and cell separation will be processed on local region or sub-image which is obtained from AML images cropping. From the evaluation results in 40 images of AML dataset, the proposed method is capable to properly segment the white blood cells region and separating the touching cell into a single cells. The average value of the segmentation results is 0.977 for precision, 0.885 for recall and 0.928 for Zijdenbos similarity index (ZSI) in white blood cell region. While in nucleus region the average value is 0.975 for precision, 0.924 for recall and 0.948 for ZSI. On cell counting, the error rate is also low which about 7.68%.

Keywords: AML, clustering, hierarchical clustering analysis, k-means, segmentation

I. PENDAHULUAN

KELAINAN pada darah dianggap sebagai salah satu penyakit yang paling berbahaya yang dapat menyebabkan kematian dan banyak dari penyakit pada darah berhubungan dengan sel darah putih seperti Leukimia [1]. Terdapat dua jenis leukemia yaitu akut dan kronis. Salah satu jenis leukemia akut menurut klasifikasi French-American-British (FAB) yang akan dibahas pada penelitian ini adalah Acute Myeloid Leukemia (AML). AML merupakan jenis kanker yang ditandai dengan infiltrasi pada *bone marrow*, darah dan jaringan lainnya oleh

sel *proliferative*, klonal dan tidak normal dari sistem *hematopoietic* [2]. Pertumbuhan sel darah putih yang tidak normal tersebut mempengaruhi produksi sel darah normal pada *bone marrow*. Meskipun AML termasuk penyakit yang relatif jarang, di Amerika Serikat kasus ini memiliki persentase sebesar 1,2% untuk kasus kematian yang disebabkan oleh kanker dan diprediksi akan meningkat seiring dengan pertambahan umur penduduk [3].

Observasi dari apusan darah melalui mikroskop memberikan informasi kualitatif dan kuantitatif yang penting untuk membantu diagnosis dari berbagai penyakit termasuk Leukimia [4]. Identifikasi dan klasifikasi dari sel darah putih ke beberapa sub-jenis sangat bernilai penting di pengujian klinis dan laboratorium, termasuk untuk diagnosis leukemia tipe AML. Akan tetapi observasi yang dilakukan oleh ahli secara manual memiliki keterbatasan dari faktor teknis dan kelelahan yang dapat menyebabkan efisiensi dan kualitas diagnosis kemungkinan menurun sehingga diperlukan suatu sistem otomatis yang dapat membantu ahli dalam melakukan diagnosis dan mempersingkat waktu pemeriksaan [5].

Pengenalan sel darah putih secara otomatis pada pengolahan citra hematologi pada umumnya terdiri dari 4 tahapan utama yaitu pra-proses, segmentasi, ekstraksi fitur dan klasifikasi. Tahap segmentasi menjadi sangat penting karena akurasi pada tahap ekstraksi fitur dan klasifikasi bergantung pada seberapa baik hasil segmentasi sel darah putih [6]. Secara umum pada citra mikroskopis sediaan apus darah tepi terdapat tiga bagian, yaitu sel darah merah, sel darah putih dan latar belakang. Pada sel darah putih sendiri terdapat dua wilayah (*region*) utama, yaitu nukleus atau inti sel dan sitoplasma [7]. Pada tahap segmentasi citra *region* sel darah putih akan dipisahkan dengan *region* lainnya dan kemudian dilanjutkan dengan pemisahan antara *region* nukleus dan sitoplasma.

Penelitian yang fokus pada segmentasi sel darah putih dengan pendekatan atau metode yang berbeda-beda telah dilakukan sebelumnya. Beberapa penelitian menggunakan citra apus darah normal [8], [9], [10], [5], [11] dan juga citra apus darah tepi leukemia [12], [13], [14].

Metode berbasis *clustering* untuk segmentasi sel darah putih telah dilakukan pada penelitian [9], [10] [13]. Pada [9] dan [13], ROI dari citra mikroskopis dibagi menjadi tiga *cluster* yaitu sel darah putih, sel darah merah dan latar belakang. Sementara itu penelitian [10] membagi ROI citra mikroskopis ke dalam enam *cluster* yaitu tiga *cluster* untuk variasi dari warna nukleus, satu *cluster* untuk sitoplasma dan dua *cluster* untuk variasi warna dari sel darah merah. Pada metode berbasis *clustering* informasi warna dan spasial yang digunakan dan penentuan jumlah *cluster* yang tepat sangat menentukan untuk mendapatkan hasil segmentasi sel darah putih yang baik, namun metode berbasis *clustering* merupakan metode yang efektif untuk dapat melakukan segmentasi pada sel darah putih dengan citra mikroskopis yang memiliki variasi *staining* yang beragam.

Permasalahan lain yang berpengaruh dalam proses analisis sel darah putih adalah keberadaan sel yang bersentuhan (*touching cell*) sehingga sel-sel tersebut perlu dideteksi dan dipisahkan [14]. Proses identifikasi dan klasifikasi sel darah putih memerlukan informasi dari tiap sel (*single cell*) dan untuk mendapatkan informasi tersebut sel yang bersentuhan harus dipisahkan dengan baik agar dapat meningkatkan akurasi pada tahap tersebut. Selain itu perhitungan jumlah sel darah putih juga penting dalam proses diagnosis ataupun tindakan medis yang dilakukan, oleh karena itu pemisahan sel yang bersentuhan yang baik akan meningkatkan akurasi perhitungan sel darah putih.

Penelitian [15] memperkenalkan metode *condition-erosion* (CE) watershed yang digunakan pada citra mikroskopik berflourensi dan merupakan pengembangan dari marker controlled watershed. Metode ini dimulai dengan proses erosi dengan batas nilai ambang (*threshold*) tertentu dan akan menjadi *extracted marker* sebagai pengganti *region minimum* pada watershed. Pada penelitian [16] yang juga menggunakan watershed, diusulkan metode *active contour* sebagai inisialisasi awal segmentasi dan *shape marker* sebagai *marker extraction* didapatkan melalui ekstraksi *adaptive H-minima* transform. Untuk memisahkan *region* pada watershed digunakan *marking function* yang dibentuk dengan fungsi jarak *euclidean distance*. Namun pada kedua metode tersebut jika *marker* tidak merepresentasikan objek dengan baik akan mengakibatkan terjadinya *under segmentation*.

Pada sel darah putih, penelitian yang menggunakan watershed untuk memisahkan sel yang bersentuhan diusulkan pada penelitian [14]. Citra sel darah putih diidentifikasi untuk penentuan sel yang bersentuhan atau tidak menggunakan analisis pada kebulatan (*roundness*) sel darah putih. Sel darah putih yang memiliki nilai kebulatan di bawah *threshold* akan dianggap sebagai sel yang berkumpul atau bersentuhan dan kemudian dipisahkan menggunakan metode watershed dengan perhitungan *distance transform*. Kontur hasil pemisahan sel darah putih dihaluskan dengan *line refining* dari ekstraksi pada transformasi watershed. Hasil pemisahan sel menunjukkan hasil yang baik pada jumlah tertentu, namun pada citra sel darah putih yang memiliki sel bersentuhan yang cukup banyak terjadi penurunan pada akurasi segmentasi.

Pendekatan lain yang berbasis *clustering* untuk pemisahan sel yang bersentuhan juga diusulkan [17], [18]. Pada penelitian [17] pemisahan dari sel yang bertumpuk pada citra sel serviks diformulasikan dengan analisis *cluster*.

Cluster validity index digunakan sebagai parameter untuk mengetahui jumlah sel yang *overlap* yang kemudian *cluster* dianalisis menggunakan Unsupervised Bayesian Classifier. Sementara itu penelitian [18] mengusulkan metode pemisahan pada citra sel darah merah menggunakan metode K-Means. Citra sel darah merah pada tahap awal di segmentasi menggunakan operasi morfologi dan untuk menentukan keberadaan dari sel darah merah yang bertumpuk pada citra digunakan fitur geometri. Metode berbasis *clustering* yang diusulkan tersebut berhasil memisahkan sel dengan lebih baik dibanding metode lainnya [18], namun jumlah *cluster* masih belum dapat ditentukan secara langsung sehingga harus ditentukan terlebih dahulu sebelum proses analisis *cluster* dan analisis *cluster* hanya terbatas pada fitur ruang spasial.

Oleh karena itu, pada penelitian ini diusulkan metode segmentasi dan pemisahan sel darah putih bersentuhan pada citra AML menggunakan analisis *cluster* dengan metode K-Means dan Hierarchical Clustering Analysis (HCA). Segmentasi akan dilakukan pada masing-masing *region* lokal atau sub-citra untuk menyeimbangkan komposisi piksel dari tiap *region* yang ada pada citra. Dalam analisis *cluster* digunakan informasi warna yaitu komponen S pada ruang warna HSV untuk mengatasi efek iluminasi akibat pencahayaan yang kurang baik. Dari hasil segmentasi, sel yang bersentuhan dipisahkan dengan metode HCA untuk mendapatkan hasil *cluster* yang stabil dengan menggunakan fitur spasial dan *gradient magnitude*.

II. DATA DAN USULAN METODE

Pada tahap ini akan dibahas data yang digunakan dan kemudian dilanjutkan dengan pembahasan detail metode yang diusulkan.

A. Data

Dataset yang digunakan pada penelitian ini adalah data citra Acute Myeloid Leukemia (AML), yang diambil di Laboratorium Kesehatan Pemerintah Daerah Kalimantan Selatan. Dataset citra AML merupakan citra sampel sediaan apus darah tepi (*peripheral blood smear*) yang di dalamnya terdapat sel darah putih normal dan abnormal (*blast*) dari pasien penderita AML dan yang telah mendapat izin untuk digunakan oleh ahli patologi dan pihak terkait untuk kepentingan ilmu pengetahuan dan pendidikan. Dataset citra AML dapat digunakan untuk pengujian kemampuan algoritma segmentasi dan sistem klasifikasi karena terdiri dari citra dengan proses *staining*, resolusi, dan pencahayaan yang berbeda-beda.

B. Metode yang Diusulkan

Pada penelitian ini diusulkan metode segmentasi dan pemisahan *touching cell* menggunakan K-Means dan *Hierarchical clustering analysis* pada citra AML. Terdapat beberapa tahapan utama pada metode yang diusulkan yaitu penentuan *initial region of interest* (ROI) sel darah putih, *cropping* citra, segmentasi sub-citra sel darah putih dengan metode K-Means, identifikasi *touching cell*, pemisahan sel dengan *hierarchical clustering analysis* dan terakhir tahap pembersihan citra. Skema utama dari metode yang diusulkan dapat dilihat pada Gambar 1.

1) Penentuan Initial ROI

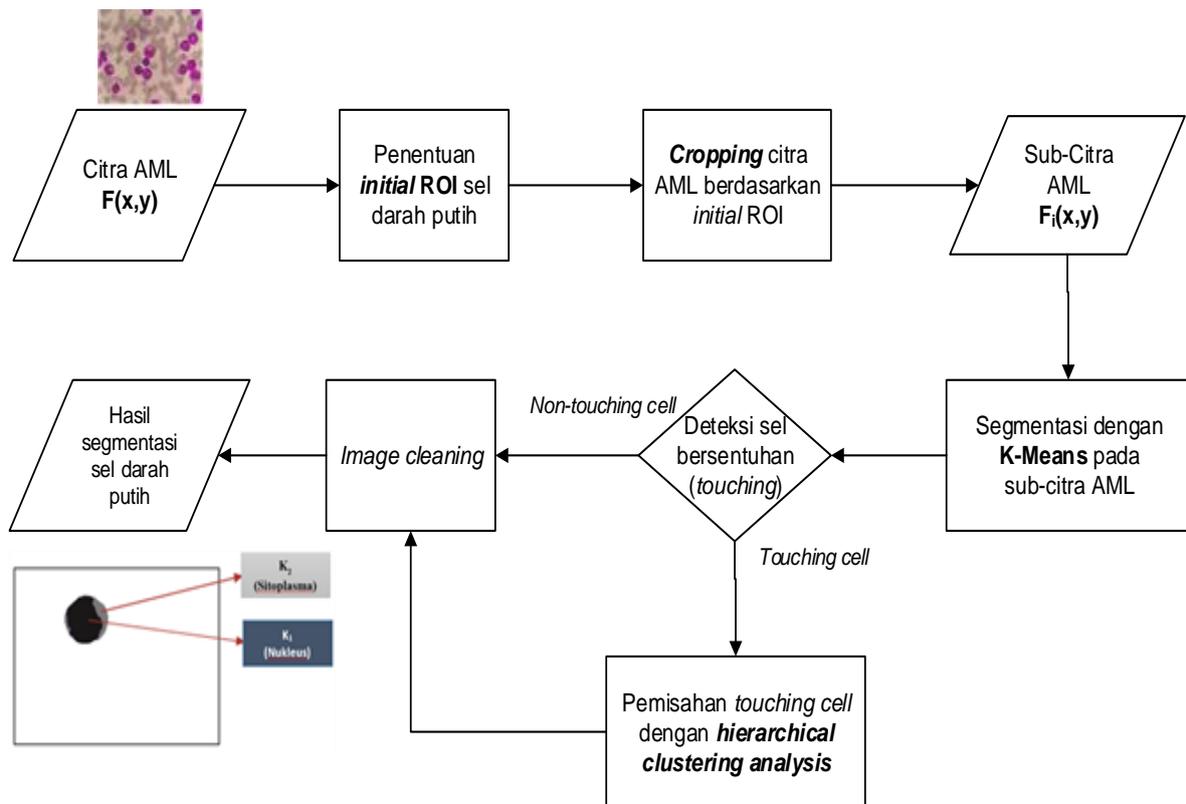
Pada tahap penentuan *initial ROI*, nukleus (inti sel) dari sel darah putih menjadi *region* yang menjadi dasar perkiraan lokasi awal dari ROI karena nukleus umumnya memiliki intensitas warna piksel lebih gelap dibandingkan *region* lainnya. *Region* dari nukleus didapatkan dengan menggunakan metode *thresholding* Otsu pada citra AML yang telah di pra-proses dengan Leukocytes Nucleus Enhancer (LNE) yang digunakan untuk memperkuat *region* nukleus [11].

Dua komponen digunakan pada pra-proses dengan LNE yaitu komponen G dari ruang warna RGB dan S pada ruang warna HSV. Secara umum perbedaan komponen S dan G pada sel darah merah kecil namun besar pada nukleus dari sel darah putih. Pada LNE nilai dari komponen S dan G dinormalisasi terlebih dahulu menggunakan persamaan (1) dan citra LNE diperoleh melalui persamaan (2).

$$IG = (I_G - G_m)/(G_M - G_m) \quad (1)$$

$$IS = (I_S - S_m)/(S_M - S_m) \quad (2)$$

$$I_E = \frac{I_S}{I_G} \quad (3)$$



Gambar. 1 Diagram alir dari metode yang diusulkan

2) Cropping Citra AML

Berdasarkan *initial ROI* yang diperoleh pada tahap sebelumnya, dilakukan *cropping* citra yang bertujuan membagi *input* citra AML menjadi beberapa sub-citra atau citra lokal yang memiliki *region* sel darah putih. Proses pembagian *region* sel darah putih menjadi sub-citra juga berguna untuk membuat proses analisis *cluster* pada tahap segmentasi menjadi lebih efektif karena rasio jumlah piksel antara *region* sel darah putih dengan *background* menjadi lebih seimbang.

Citra biner *initial ROI* sel darah putih adalah $r(x,y)$, maka $r_i(x,y)$ adalah komponen terhubung (*region-region*) *initial ROI* sel darah putih, dengan $i=1, 2, \dots, n$; n adalah jumlah komponen terhubung *initial ROI* sel darah putih dan $r_i(x,y) \in r(x,y)$. *Bounding box* adalah kotak imajiner yang mengelilingi $r_i(x,y)$. *Bounding box* awal tiap $r_i(x,y)$ ditentukan berdasarkan koordinat spasial piksel *upper-left* (UL), *upper-right* (UR), *lower-left* (LL), dan *lower-right* (LR) dari $r_i(x,y)$ [19].

3) Segmentasi Sel Darah Putih dengan Metode K-Means

Dari sub-citra hasil *cropping* pada citra AML, segmentasi sel darah putih dilakukan dengan analisis *cluster* menggunakan metode K-means. Dataset citra AML yang digunakan pada penelitian ini memiliki variasi *staining* yang beragam sehingga analisis berbasis *cluster* yang bersifat adaptif terhadap variasi informasi piksel pada dataset citra AML lebih efektif untuk mendapatkan *region* dari sel darah putih.

Metode K-Means mengelompokkan data secara iteratif dengan melakukan partisi set data ke sejumlah K *cluster* yang telah ditentukan di awal [20]. Metode K-Means melakukan partisi pada informasi warna pada tiap piksel dan dapat mengatasi variasi informasi dataset yang memiliki variasi *staining* beragam. Data yang digunakan pada proses analisis *cluster* pada citra AML adalah komponen *saturation* (S) pada ruang warna HSV untuk mengatasi efek iluminasi akibat dari pencahayaan yang kurang baik.

Sub-citra AML, $F'(x,y)$ pada analisis *cluster* menggunakan metode K-Means merupakan titik data $X = \{x_i \mid i=1, \dots, N\}$ yang dipartisi ke dalam K *cluster*. Pada proses segmentasi, informasi warna piksel citra AML, $F'(x,y)$, dikelompokkan menjadi tiga *cluster* yaitu K_1 merupakan *region* nukleus, K_2 merupakan *region* sitoplasma dan K_3 merupakan *region background* (sel darah merah dan *platelets*). Untuk mengukur ketidakmiripan (*dissimilarity*)

atau jarak dalam penentuan keanggotaan data ke dalam satu *cluster* digunakan fungsi jarak *Euclidean distance* yang dituliskan pada persamaan 4.

$$d(x, c) = \|x - c\| = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}, \quad i = 1, 2, \dots, n \quad (4)$$

Dari *cluster* K_1 dan K_2 yang merupakan *region* nukleus dan sitoplasma ditentukan sebagai *region* sel darah putih, $f_{bin}(x,y)$. Pada sub-citra biner hasil segmentasi sel darah putih dengan metode K-Means akan menjadi dasar untuk pemisahan sel pada tahap selanjutnya.

4) Identifikasi Touching Cell

Dari setiap sub-citra yang telah di segmentasi dengan metode K-Means dilakukan identifikasi untuk menentukan sub citra sel darah putih yang memiliki *region touching cell* dan sel darah putih tunggal. Untuk mendeteksi *region* yang termasuk *touching cell* digunakan fitur geometri yaitu kebulatan (*roundness*) dan luasan area (*size*). Pada analisis pada fitur geometri dengan menghitung kebulatan sel, secara umum sel darah putih tunggal memiliki bentuk lebih bundar dibanding sel yang bersentuhan. Perhitungan kebulatan dari *region* sel didapatkan dengan berdasarkan persamaan 5.

$$roundness = \frac{4 \cdot \pi \cdot area}{perimeter^2} \quad (5)$$

Roundness menyatakan *circularity* objek, yaitu rasio antara *area* dengan *perimeter*. *Roundness* bernilai satu untuk objek yang berbentuk bundar, sedangkan bernilai kurang dari satu untuk objek yang menjauhi bentuk bundar. Pada penelitian [14] *roundness threshold* yang diusulkan sebesar 0.8.

Untuk deteksi dengan fitur *roundness* terdapat kemungkinan beberapa sel yang berkumpul mendekati bentuk bundar sehingga fitur luasan area (*size*) diperlukan untuk meningkatkan akurasi dari pendeteksian *touching cell*. Area atau jumlah piksel pada sel darah putih yang bersentuhan relatif berukuran lebih besar dibanding rata-rata ukuran sel tunggal sehingga sel dengan ukuran piksel diatas nilai *threshold* sel tunggal akan diklasifikasikan sebagai *touching cell*. *Threshold* untuk luasan area ditentukan dengan menghitung rata-rata ukuran sel tunggal. Untuk nilai dari rata-rata luas dari sel darah putih tunggal diperoleh melalui perhitungan pada persamaan 6.

$$single\ cell\ mean = mean(area_{wbc}) - (0,25 \cdot std(area_{wbc})) \quad (6)$$

5) Pemisahan Touching Cell dengan Hierarchical Clustering Analysis

Pada proses ini, sub-citra yang ditentukan sebagai *touching cell* pada perhitungan dengan fitur geometri sebelumnya akan dipisahkan dengan menggunakan metode Hierarchical Clustering Analysis. Pemisahan sel darah putih yang bersentuhan dilakukan dengan proses klasterisasi dengan ruang informasi spasial (piksel) dan *gradient magnitude* dari citra AML. Setiap koordinat piksel pada *region* sel darah putih merupakan titik yang merupakan informasi spasial dari sub-citra $f_i(x,y)$. Sedangkan informasi *gradient* yaitu sub-citra $f_G(x,y)$ berfungsi untuk memberikan informasi seberapa besar (*magnitude*) perubahan intensitas yang terjadi pada sub-citra di *region* sel darah putih.

Hierarchical clustering analysis merupakan metode analisis *cluster* yang memberikan hasil *cluster* yang stabil. Metode pengelompokan pada hirarki *cluster* yang digunakan pada penelitian ini adalah aglomeratif yang memulai pengelompokan dari masing-masing data sebagai *cluster* kemudian secara rekursif mencari *cluster* terdekat sebagai pasangan untuk digabungkan sebagai satu *cluster* yang lebih besar. Kedekatan dua *cluster* pada penelitian ini dihitung dengan metode *centroid linkage* yaitu menggabung *cluster* berdasarkan jarak terdekat antara titik pusat dari dua *cluster*. Jarak dua *cluster* pada metode *centroid linkage* diukur dengan dengan formula jarak tengah (*median proximity*) dengan persamaan 7.

$$d(U, V) = \|C_u - C_v\|; d(U, V) \in D \quad (7)$$

Dimana D menyatakan matriks kedekatan (jarak) antar *cluster*. Untuk parameter perhitungan jarak yang digunakan adalah *Euclidean distance* seperti yang dituliskan pada persamaan 4. Penerapan *hierarchical clustering analysis* pada sub-citra juga mengoptimalkan proses analisis *cluster* karena pada citra dengan area piksel lebih kecil

proses klusterisasi menjadi lebih cepat dibandingkan pada citra AML secara keseluruhan. Untuk kernel yang digunakan pada fitur *gradient magnitude* adalah operator *Roberts*.

6) Pembersihan Citra

Pembersihan citra (*image cleaning*) terdiri dari penghapusan semua region sel darah putih yang terdapat pada tepi citra dan semua komponen abnormal (*region* sel darah putih cacat pewarnaan), bertujuan untuk mengurangi *error* pada tahap klasifikasi. Penghapusan citra yang berada di tepi merupakan operasi sederhana, sedangkan penghapusan komponen abnormal merupakan proses yang lebih kompleks karena membutuhkan analisis geometri *region* sel darah putih.

Pertama, ukuran *area* dan *convex area* dihitung untuk masing-masing sel darah putih yang ada pada citra. Selanjutnya kombinasi *area* dan *convex area* digunakan untuk menghitung nilai *solidity* yang digunakan untuk memisahkan komponen abnormal. *Solidity* mengukur kepadatan objek. *Solidity* didefinisikan sebagai rasio antara objek *area* dengan *convex hull* objek seperti pada persamaan 8.

$$solidity = \frac{area}{convex\ area} \quad (8)$$

Nilai *solidity* 1 menandakan objek yang solid, dan nilai kurang dari 1 menandakan objek dengan *boundary irregular* (atau terdapat *holes*). Nilai ambang *solidity* yang digunakan pada penelitian ini yaitu 0.9 [14]. Semua objek bernilai *solidity* di bawah nilai ambang (*threshold*) akan dihapus.

III. HASIL DAN PEMBAHASAN

Uji coba pada penelitian ini menggunakan data 40 citra AML dari basis data yang ada di Laboratorium Kesehatan Pemerintah Daerah Kalimantan Selatan. Citra uji tersebut merupakan citra sediaan apus darah tepi (*peripheral blood smear*) sel darah yang mengandung sel-sel AML (*blast*) dan sel-sel normal (*healthy cells*). Beberapa citra pada dataset terdapat sejumlah *region* sel yang berkelompok atau *touching cell*.

Hasil segmentasi dari metode yang diusulkan (analisis *cluster* dengan metode K-Means pada citra lokal) dibandingkan dengan citra yang telah disegmentasi secara manual. Citra yang disegmentasi secara manual tersebut dianggap sebagai hasil segmentasi yang benar (*ground truth*). Perhitungan manual jumlah sel darah putih pada data berdasar pada arahan yang diberikan oleh ahli/dokter patologi.

Untuk Hasil segmentasi metode yang diusulkan dan citra *ground truth* digunakan untuk mengklasifikasikan piksel ROI sel darah putih sebagai *true positive*, *false negative*, atau *false positive*, yang direpresentasikan ke dalam *Zijdenbos Similarity Index* (ZSI), presisi, dan *recall* [19]. ZSI didefinisikan sebagai rasio dua kali interseksi area antara hasil segmentasi dan citra *ground truth* dengan jumlah dari keduanya. Nilai ZSI dihitung menggunakan persamaan 9.

$$ZSI = \frac{2 \times |A \cap G|}{|A| + |G|} \quad (9)$$

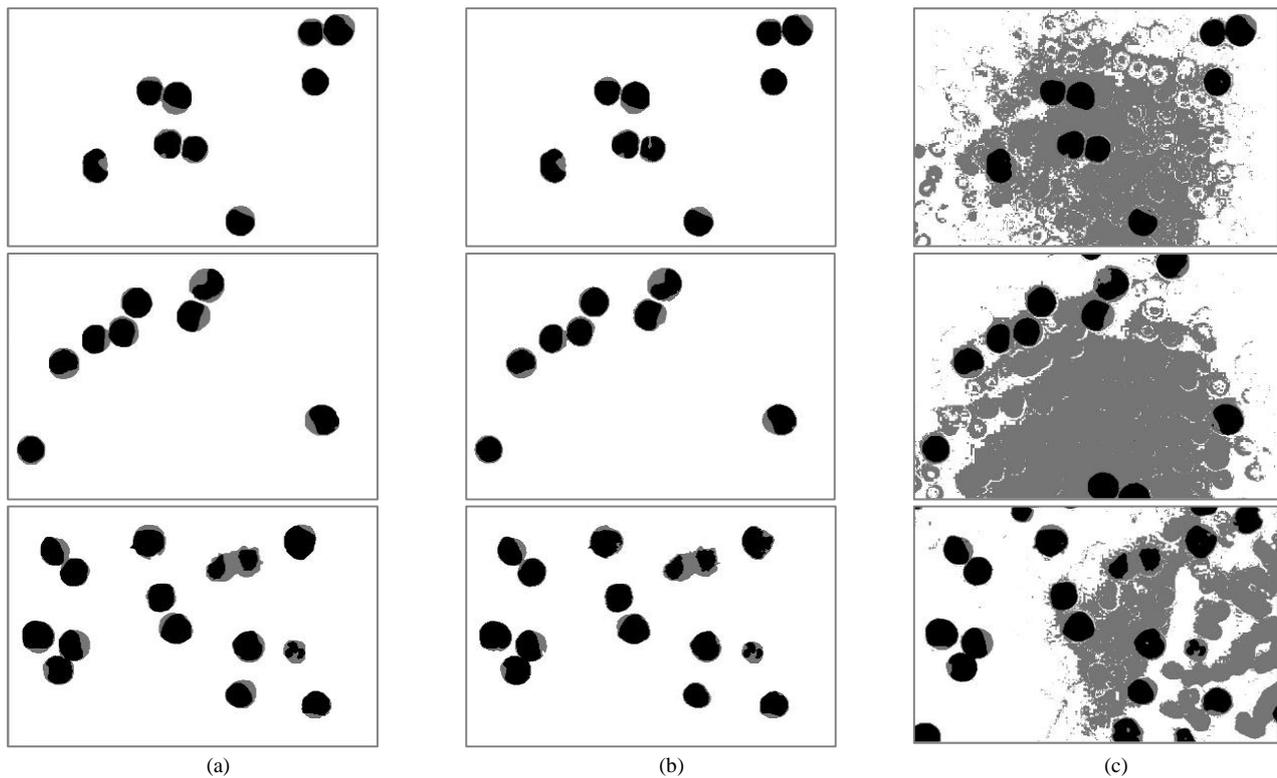
Dengan *A* merepresentasikan piksel hasil segmentasi metode yang diusulkan dan *G* merepresentasikan piksel citra *ground truth*. Dengan menggunakan *true positive* (TP), *false negative* (FN), dan *false positive* (FP), presisi (P) dan *recall* (R) dapat dihitung menggunakan persamaan 10 dan 11.

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

Uji coba pada penelitian ini dilakukan pada PC dengan spesifikasi CPU i3 2.70 GHz dan RAM 4GB. Pada tahap uji coba segmentasi terhadap *output* dari metode yang diusulkan, terdapat dua skenario uji coba yang dilakukan yaitu perbandingan hasil evaluasi segmentasi antara metode yang diusulkan (LKM) dengan analisis *cluster* menggunakan metode K-Means pada citra global (GKM) terhadap *region* sel darah putih dan nukleus. Untuk mengevaluasi hasil segmentasi citra AML digunakan nilai dari presisi, *recall*, dan *Zijdenbos similarity index* (ZSI).

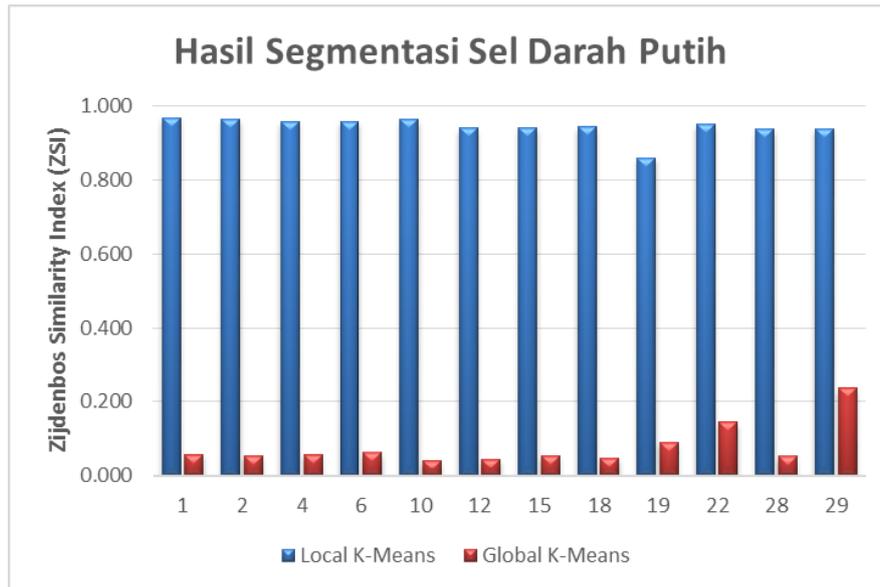
Citra *ground truth* segmentasi manual dari citra AML dibandingkan dengan hasil segmentasi untuk mendapatkan nilai presisi, *recall*, dan ZSI.



Gambar. 1 Perbandingan hasil segmentasi citra AML dengan citra *ground-truth*.
(a) *Gradient magnitude*, (b) *Local K-Means* dan (c) *Global K-Means*

LKM dan GKM melakukan analisis *cluster* terhadap nilai *saturation* (S) pada ruang warna *Hue Saturation Value* (HSV). Pada GKM, analisis *cluster* dilakukan pada nilai *saturation* dari keseluruhan citra. Pembagian jumlah *cluster* yang dilakukan pada GKM sama seperti LKM, yaitu hasil analisis *cluster* dibagi menjadi *region* nukleus, sitoplasma dan *background*. Gambar 1 menunjukkan perbandingan antara hasil segmentasi metode LKM dengan GKM. *Region* dari nukleus ditunjukkan dengan warna hitam, sedangkan sitoplasma dengan warna keabuan.

Pada Tabel I ditunjukkan evaluasi hasil segmentasi pada *region* sel darah putih menggunakan LKM dan GKM dengan nilai presisi (P), *recall* (R) dan *Zijdenbos similarity index* (ZSI) pada citra uji dataset AML. Untuk perbandingan hasil evaluasi segmentasi sel darah putih antara metode LKM dan GKM ditunjukkan oleh grafik pada Gambar 2.

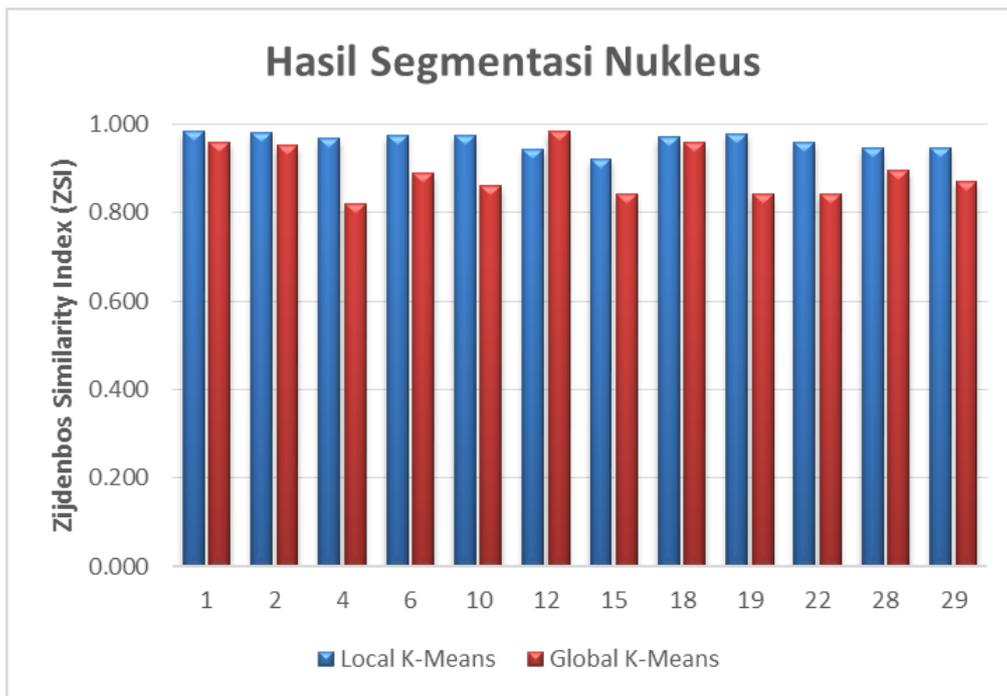


Gambar. 2 Evaluasi hasil segmentasi sel darah putih berdasarkan ZSI.

Tabel I
EVALUASI HASIL SEGMENTASI SEL DARAH PUTIH

Citra	LKM			GKM		
	P	R	ZSI	P	R	ZSI
aml_01	0.997	0.938	0.967	0.030	0.988	0.057
aml_02	0.996	0.935	0.964	0.028	0.996	0.054
aml_03	0.996	0.934	0.964	0.026	0.991	0.051
aml_04	0.992	0.927	0.958	0.029	0.974	0.057
aml_05	0.993	0.932	0.962	0.040	0.964	0.077
aml_06	0.997	0.919	0.957	0.033	0.935	0.063
aml_07	0.996	0.906	0.949	0.046	0.927	0.087
aml_08	0.998	0.935	0.966	0.022	0.998	0.044
aml_09	0.999	0.917	0.956	0.032	0.951	0.061
aml_10	1.000	0.928	0.963	0.021	1.000	0.042
aml_12	0.999	0.888	0.940	0.023	0.906	0.045
aml_13	0.998	0.902	0.948	0.023	0.936	0.044
aml_14	0.998	0.904	0.949	0.016	0.970	0.031
aml_15	0.998	0.890	0.941	0.028	0.843	0.055
aml_16	0.998	0.872	0.931	0.051	0.883	0.097
aml_18	0.987	0.902	0.943	0.025	0.995	0.048
aml_19	0.993	0.757	0.859	0.047	0.970	0.090
aml_22	0.997	0.908	0.950	0.078	0.964	0.144
aml_28	0.999	0.881	0.937	0.027	0.979	0.052
aml_29	0.999	0.884	0.938	0.149	0.569	0.236

Untuk region nukleus (inti sel), evaluasi hasil segmentasi antara LKM dan GKM ditunjukkan oleh Tabel II dengan nilai presisi (P), recall (R) dan Zidenbos similarity index (ZSI). Untuk perbandingan dari hasil evaluasi segmentasi pada region nukleus diperlihatkan oleh grafik pada Gambar 3.



Gambar. 3 Evaluasi hasil segmentasi nukleus berdasarkan ZSI.

Tabel II
EVALUASI HASIL SEGMENTASI NUKLEUS

Citra	LKM			GKM		
	P	R	ZSI	P	R	ZSI
aml_01	0.991	0.976	0.983	0.924	0.999	0.960
aml_02	0.982	0.982	0.982	0.907	1.000	0.951
aml_03	0.992	0.969	0.980	0.927	1.000	0.962
aml_04	0.996	0.940	0.967	0.703	0.982	0.819
aml_05	0.987	0.955	0.971	0.866	0.995	0.926
aml_06	0.994	0.955	0.974	0.800	0.999	0.888
aml_07	0.989	0.953	0.970	0.845	0.997	0.915
aml_08	0.998	0.924	0.960	0.805	0.995	0.890
aml_09	0.996	0.951	0.973	0.887	0.991	0.936
aml_10	0.996	0.955	0.975	0.757	1.000	0.861
aml_12	1.000	0.892	0.943	0.988	0.979	0.983
aml_13	0.999	0.926	0.961	0.884	0.996	0.937
aml_14	0.998	0.966	0.981	0.824	0.997	0.902
aml_15	0.999	0.856	0.922	0.749	0.964	0.843
aml_16	1.000	0.862	0.926	0.872	0.979	0.923
aml_18	0.995	0.946	0.970	0.922	0.997	0.958
aml_19	0.993	0.966	0.979	0.731	0.995	0.843
aml_22	0.966	0.949	0.957	0.734	0.988	0.842
aml_28	0.997	0.897	0.945	0.823	0.983	0.896
aml_29	0.987	0.911	0.947	0.833	0.907	0.869

Pada hasil uji coba hasil segmentasi sel darah putih menggunakan metode yang diusulkan memiliki nilai rata-rata presisi lebih tinggi dibandingkan dengan hasil segmentasi sel darah putih menggunakan analisis *cluster* dengan metode K-Means pada citra global (GKM), yaitu dengan nilai sebesar 0.977 sedangkan GKM hanya memiliki nilai rata-rata presisi sebesar 0.059. Begitu juga dengan nilai rata-rata perhitungan *Zijdenbos Similarity Index (ZSI)*, metode yang diusulkan pada penelitian ini memiliki nilai rata-rata yang lebih baik dibandingkan dengan GKM yaitu sebesar 0.928, lebih tinggi dibandingkan rata-rata nilai ZSI dari GKM yang hanya 0.102. Namun pada hasil

perhitungan nilai rata-rata *recall*, hasil segmentasi dengan GKM memiliki nilai yang tidak jauh berbeda yaitu sebesar 0.796, sementara metode yang diusulkan memiliki nilai rata-rata 0.885.

Pada hasil segmentasi di *region* nukleus (inti sel) dari sel darah putih, metode yang diusulkan juga memiliki nilai rata-rata yang lebih baik pada hasil perhitungan presisi dan ZSI. Nilai rata-rata presisi dan ZSI masing-masing adalah 0.975 dan 0.948. Sedangkan untuk nilai rata-rata *recall*, hasil segmentasi oleh GKM memiliki nilai rata-rata yang sedikit lebih baik yaitu 0.958, dibanding dengan metode yang diusulkan yang memiliki nilai rata-rata *recall* sebesar 0.924. Secara umum hasil segmentasi pada metode yang diusulkan maupun GKM sama-sama memiliki hasil yang lebih baik pada *region* nukleus baik dari rata-rata nilai presisi, *recall* ataupun ZSI dibandingkan dengan hasil pada *region* sel darah putih. Namun di sisi lain perbedaan nilai rata-rata yang signifikan terdapat pada hasil perhitungan presisi dan ZSI di *region* sel darah putih dimana metode yang diusulkan memiliki nilai yang jauh lebih tinggi dengan perbedaan rata-rata sekitar 0.918 untuk presisi dan 0.825 untuk ZSI.

Pada rata-rata nilai *recall* dimana hasil segmentasi GKM memiliki nilai yang lebih baik pada *region* nukleus dibanding metode yang diusulkan, namun perbedaannya tidak terlalu signifikan dengan nilai perbedaan sekitar 0.033. Sedangkan pada *region* sel darah putih metode yang diusulkan memiliki nilai *recall* yang sedikit lebih baik dibanding GKM yaitu dengan perbedaan rata-rata sebesar 0.08. Dari hasil perhitungan rata-rata nilai *recall* juga dapat diketahui bahwa pada sebagian besar data citra AML yang disegmentasi memiliki nilai *false negative* yang cukup rendah terutama pada *region* nukleus, sehingga terjadinya *under-segmentation* juga kecil. Akan tetapi untuk *region* sel darah putih rasio terjadinya *over-segmentation* cukup tinggi pada hasil segmentasi GKM karena nilai presisi yang rendah yaitu sebesar 0.059. Rasio *false positive* yang tinggi pada GKM disebabkan pada *cluster* sitoplasma terdapat banyak sel darah merah ataupun *background* yang ikut tersegmentasi, terutama pada citra yang memiliki rasio perbandingan piksel sel darah putih yang jumlahnya jauh dibawah piksel *region* sel darah merah dan *background*. Dengan banyaknya *region* sel darah merah dan *background* yang ikut tersegmentasi sebagai objek juga menyebabkan nilai rata-rata ZSI pada hasil segmentasi sel darah putih menjadi sangat rendah pada GKM.

Pada skenario uji coba untuk mengevaluasi hasil pemisahan sel dilakukan perhitungan total terhadap jumlah sel darah putih tunggal yang ada pada citra, termasuk *region touching cell* yang telah dipisahkan menggunakan metode yang diusulkan (Hierarchical Clustering Analysis). Hasil perhitungan jumlah sel dari metode yang diusulkan dibandingkan dengan perhitungan berdasarkan data *ground truth* yang ditunjukkan oleh Tabel III. Selain jumlah keseluruhan, perhitungan juga dilakukan berdasarkan beberapa kriteria yaitu jumlah *region* sel darah putih yang diidentifikasi benar yaitu *correctly counted*, jumlah *region* yang seharusnya tidak disertakan yaitu *over counted*, dan jumlah *region* yang merupakan objek tetapi tidak terhitung yaitu *not counted*. Hasil perhitungan dari Tabel III juga ditunjukkan dalam bentuk grafik oleh Gambar 4.

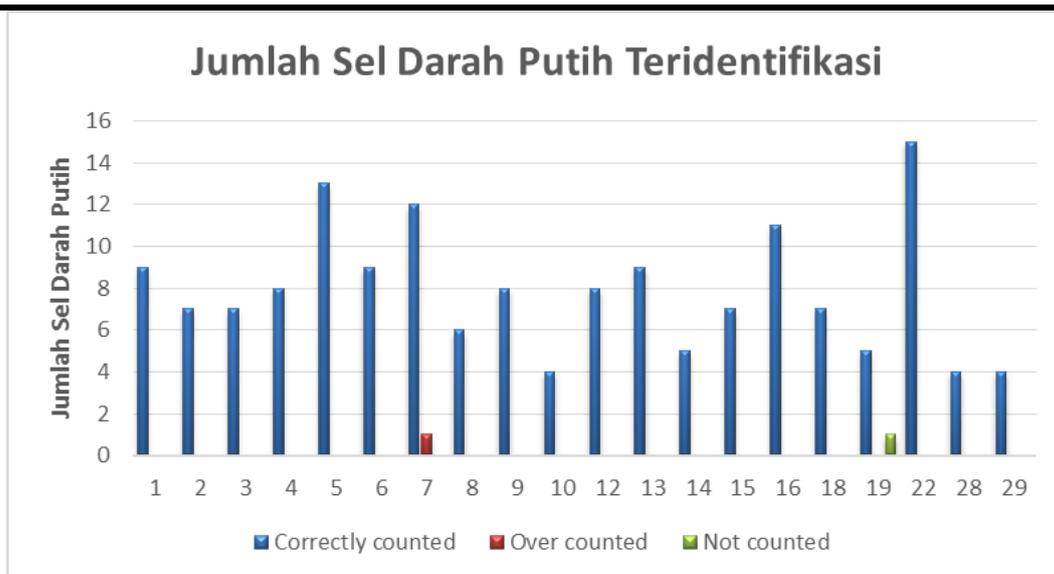
Pada ujicoba identifikasi *touching cell*, secara umum fitur geometri yang digunakan yaitu fitur *roundness* dan rata-rata luasan sel tunggal dari sel darah putih telah berhasil menemukan *region* yang merupakan *touching cell* dan juga menentukan jumlah *cluster* yang akan digunakan pada proses pemisahan sel darah putih dengan HCA. Untuk pemisahan pada *touching cell* pada citra menjadi sel-sel tunggal, penggunaan fitur ruang spasial dan *gradient magnitude* dapat memisahkan sel-sel tersebut dengan identifikasi yang tepat untuk penentuan jumlah *cluster*. Fitur ruang spasial $f'(x,y)$ merupakan fitur yang memberikan informasi letak koordinat dari sel darah putih dan fitur *gradient magnitude* memberikan informasi letak dari bagian tepi dari setiap sel tunggal. Antara *region* nukleus dan sitoplasma terdapat perbedaan intensitas yang cukup besar sehingga berguna sebagai batas pemisah.

Namun karena pada fitur *gradient magnitude* secara umum *region* sitoplasma akan menjadi *region* yang lebih terang yang kemudian menjadi pembatas untuk memisahkan sel, pada beberapa sub-citra terdapat sel darah putih yang *region* sitoplasmanya memiliki rasio yang cukup besar atau hampir sebanding dengan nukleus sehingga perubahan intensitas *gradient* letaknya menjadi jauh dari *region* yang ideal sebagai *region* pemisahan sel.

Tabel III
PERBANDINGAN ANTARA JUMLAH SEL DARAH PUTIH DENGAN PERHITUNGAN MANUAL DAN YANG TERIDENTIFIKASI MENGGUNAKAN METODE YANG DIUSULKAN

Citra	Ground truth	Metode yang diusulkan (HCA)			
		Counted	Correctly counted	Over counted	Not counted
aml_01	9	9	9	0	0
aml_02	7	7	7	0	0

aml_03	7	7	7	0	0
aml_04	8	8	8	0	0
aml_05	13	13	13	0	0
aml_06	9	9	9	0	0
aml_07	12	13	12	1	0
aml_08	6	6	6	0	0
aml_09	8	8	8	0	0
aml_10	4	4	4	0	0
aml_12	8	8	8	0	0
aml_13	9	9	9	0	0
aml_14	5	5	5	0	0
aml_15	7	7	7	0	0
aml_16	11	11	11	0	0
aml_18	7	7	7	0	0
aml_19	6	5	5	0	1
aml_22	15	15	15	0	0
aml_28	4	4	4	0	0
aml_29	4	4	4	0	0



Gambar. 4 Jumlah sel darah putih teridentifikasi.

Selain kemungkinan luas *region* sitoplasma yang cukup besar sehingga menyebabkan hasil pemisahan tidak baik, letak nukleus yang saling berdekatan langsung juga dapat mengganggu proses pemisahan sel. Hal ini karena nukleus yang berdekatan memiliki intensitas yang mirip sehingga perubahan *gradient* menjadi kecil (memiliki nilai yang rendah) pada *region* yang seharusnya menjadi *region* ideal pemisahan sel. Karena terdapat *region* yang batasnya tidak jelas tersebut, saat proses analisis *cluster* dengan HCA dapat terjadi *over-cluster*.

IV. KESIMPULAN

Pada penelitian ini telah dilakukan segmentasi dan pemisahan sel darah putih menggunakan metode K-Means dan *hierarchical clustering analysis* pada citra AML. Analisis *cluster* pada citra local (LKM) menghasilkan segmentasi sel darah putih dan nukleus lebih akurat dengan rata-rata ZSI 82% dan 6.1% lebih tinggi dari pada menggunakan analisis *cluster* metode K-Means pada citra global (GKM). Metode yang diusulkan memiliki kelebihan dengan memiliki komposisi jumlah piksel *region* sel darah putih yang lebih seimbang antara *region* sel darah putih dengan *background* saat proses analisis *cluster* sehingga dapat melakukan segmentasi lebih baik terutama pada *region* sitoplasma. Pada proses identifikasi dan pemisahan *touching cell* pada sub-citra, fitur

geometri *roundness* dan rata-rata luasan sel tunggal yang kemudian dipisahkan dengan metode *hierarchical clustering analysis* secara umum telah dapat menentukan *region* dari *touching cell* dengan baik dan memisahkannya menjadi sel-sel tunggal.

Penelitian selanjutnya yang dapat dikembangkan dari metode yang diusulkan adalah dengan menemukan komposisi ruang warna yang mampu membuat *region* sitoplasma menjadi lebih berbeda dan memilih fitur yang lebih baik untuk proses pemisahan *touching cell* dengan *hierarchical clustering analysis* atau metode *post-processing* yang akan memperbaiki hasil pemisahan sel yang kurang baik.

DAFTAR PUSTAKA

- [1] H.T. Madhloom, S.A. Kareem, H. Ariffin, A.A. Zaidan, H.O. Alanazi and B.B. Zaidan, 2010. An Automated White Blood Cell Nucleus Localization and Segmentation using Image Arithmetic and Automatic Threshold. *Journal of Applied Sciences*, 10: 959-966.
- [2] H. Döhner, D.J. Weisdorf, C.D. Bloomfield (17 September 2015). "Acute Myeloid Leukemia". *The New England Journal of Medicine*. 373 (12): 1136–52.
- [3] A. Jemal, A. Thomas, T. Murray, M. Thun (2002). "Cancer statistics, 2002". *CA Cancer J Clin*. 52 (1): 23–47.
- [4] Theera-Umporn, N. Wang, L. & Jin, Y. (Eds.) White Blood Cell Segmentation and Classification in Microscopic Bone Marrow Images Fuzzy Systems and Knowledge Discovery: *Second International Conference, FSKD 2005*, Changsha, China, August 27-29, 2005, Proceedings, Part II, Springer Berlin Heidelberg, 2005, 787-796.
- [5] S. H. Rezaatofighi, H. Soltanian-Zadeh, Automatic recognition of five types of white blood cells in peripheral blood, *Computerized Medical Imaging and Graphics*, Volume 35, Issue 4, June 2011, Pages 333-343, ISSN 0895-6111.
- [6] R.A. Saputra, C. Fatichah, N. Suciati. Penggabungan Fitur Tekstur yang Invariant terhadap Iluminasi dan Fitur Bentuk untuk Deteksi Acute Lymphoblastic Leukemia. *JBI: Jurnal Buana Informatika*, vol. 7, No. 1, 2016.
- [7] C. Fatichah, M.L. Tangel, M.R. Widyanto, F. Dong and K. Hirota. Interest-Based Ordering for Fuzzy Morphology on White Blood Cell Image Segmentation. *JACII: Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.16, No.1, pp. 76-86, 2012.
- [8] V. Piuri and F. Scotti, "Morphological classification of blood leucocytes by microscope images," *2004 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, 2004. CIMSAA, 2004, pp. 103-108.
- [9] F. Scotti, "Robust Segmentation and Measurements Techniques of White Cells in Blood Microscope Images," *2006 IEEE Instrumentation and Measurement Technology Conference Proceedings*, Sorrento, 2006, pp. 43-48.
- [10] E.-y. Wang, Z. Gou, A.-m. Miao, S.-q. Peng, Z.-y. Niu and X.-l. Shi. Recognition of Blood Cell Images Based on Color Fuzzy Clustering Fuzzy. *Information and Engineering*, Volume 2, Springer Berlin Heidelberg, 2009, pp. 69-75.
- [11] D.C. Huang, K.D. Hung, Y.K. Chan, A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images, *Journal of Systems and Software*, Volume 85, Issue 9, September 2012, Pages 2104-2118, ISSN 0164-1212.
- [12] H.T. Madhloom, S.A. Kareem and H. Ariffin. An Image Processing Application for the Localization and Segmentation of Lymphoblast Cell Using Peripheral Blood Images, *Journal of Medical Systems*, Vol. 36, pp. 2149-2158, 2012.
- [13] S. Mohapatra, D. Patra and S. Satpathy. An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images, *Neural Computing and Applications*, 24, 1887-1904, 2014.
- [14] L. Putzu, G. Caocci, C. Di Ruberto. Leucocyte classification for leukaemia detection using image processing techniques, *Artificial Intelligence in Medicine*, Volume 62, Issue 3, November 2014, Pages 179-191, ISSN 0933-3657.
- [15] X. Yang, H. Li and X. Zhou, "Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 11, pp. 2405-2414, Nov. 2006.
- [16] J. Cheng and J. C. Rajapakse. "Segmentation of Clustered Nuclei With Shape Markers and Marking Function," in *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 741-748, March 2009.
- [17] C. Jung, C. Kim, S. W. Chae and S. Oh. "Unsupervised Segmentation of Overlapped Nuclei Using Bayesian Classification," in *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2825-2832, Dec. 2010.
- [18] F. Effendy, "Segmentasi sel darah merah bertumpuk berdasarkan fitur geometri pada perhitungan sel darah merah", tesis magister, Teknik Informatika, ITS, Surabaya, Indonesia, 2013.
- [19] E.P. Mandiyarta, C. Fatichah. Three-level Local Thresholding Berbasis Metode Otsu untuk Segmentasi Leukosit pada Citra Leukemia Limfoblastik Akut. *JBI: Jurnal Buana Informatika*, vol. 7, No. 1, 2016.
- [20] E. Prasetyo, "Pengolahan Citra Digital dan Aplikasinya Menggunakan Matlab". Yogyakarta, Indonesia, 2011.