

Identifikasi Sel Darah Merah Bertumpuk Menggunakan Pohon Keputusan *Fuzzy* Berbasis Gini Index

Eka Prakarsa Mandiyartha¹, Muchammad Kurniawan², Rizal Setya Perdana³

Program Studi Teknik Informatika, Fakultas Teknologi Informasi,

Institut Teknologi Sepuluh Nopember

Kampus ITS Keputih, Sukolilo, Surabaya 60111, Jawa Timur

E-mail: ¹eka.prakarsa13@mhs.if.its.ac.id, ²muchammad.kurniawan13@mhs.if.its.ac.id,

³rizal13@mhs.if.its.ac.id

Masuk: 10 Juni 2014; Direvisi: 26 Juni 2014, 7 Juli 2014; Diterima: 8 Juli 2014

Abstract. *This paper proposes a data mining approach for the identification of red blood cells overlapping in the blood smear images to improve red blood cells counting accuracy. Geometry and color features are used. The geometry features consist of the area and eccentricity of the cell. In the identification process, fuzzy approach is applied. All features are represented by fuzzy membership function. Identification is done based on rules which is derived from the generated fuzzy decision tree using multi split branching. Gini index is used to measure the split attribute. The results of the test which has been performed on 10 blood smear images containing 532 red blood cells show that the average accuracy of the proposed method is 96,14%. With the high accuracy, it is expected that the method could improve the diagnosis accuracy of diseases based on the number of red blood cells.*

Keywords: *fuzzy decision tree, multi split decision tree, red blood cells, overlapping red blood cells*

Abstrak. *Pendekatan teknik data mining diusulkan untuk identifikasi sel darah merah bertumpuk pada citra makroskopik sel darah untuk meningkatkan akurasi penghitungan jumlah sel darah merah. Fitur yang digunakan adalah geometri dan warna. Fitur geometri terdiri dari luasan dan eksentrisitas sel. Pada proses identifikasi digunakan pendekatan fuzzy. Setiap fitur direpresentasikan dengan fungsi keanggotaan fuzzy. Identifikasi dilakukan berdasarkan aturan yang diperoleh dari pohon keputusan fuzzy yang dibangkitkan. Pencabangan multisplit digunakan pada pohon keputusan fuzzy. Pengukuran split atribut menggunakan nilai gini index. Hasil pengujian pada 10 citra makroskopik sel darah yang mengandung 532 sel darah merah menunjukkan bahwa metode yang diusulkan memiliki rata-rata akurasi sebesar 96,14%. Dengan akurasi yang tinggi diharapkan dapat meningkatkan akurasi diagnosis penyakit berdasarkan jumlah sel darah merah.*

Kata kunci: *pohon keputusan fuzzy, pohon keputusan multisplit, sel darah merah, sel darah merah bertumpuk*

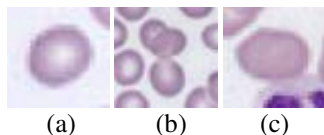
1. Pendahuluan

Sel darah merah atau disebut juga eritrosit merupakan sel darah yang jumlahnya terbanyak dalam tubuh manusia (Mahmood, 2012). Jumlah sel darah merah dapat memberikan informasi yang mengindikasikan adanya gangguan hematologi. Gangguan hematologi adalah gangguan pada pembentukan sel darah merah, meliputi penurunan dan peningkatan jumlah sel (polisitemia). Penurunan jumlah sel darah merah ditemukan pada penyakit kronis, seperti penyakit hati, anemia, dan leukemia, sedangkan polisitemia ditemukan pada penderita diare, dehidrasi berat, luka bakar, maupun pendarahan berat. Penghitungan sel darah merah dilakukan dalam proses diagnosis beberapa penyakit tersebut. Penghitungan sel darah merah di laboratorium dapat dilakukan secara manual, menggunakan *hemocytometer* dan mikroskop, atau

menggunakan mesin *hematology analyzer*. Penghitungan secara manual melibatkan proses laboratorium yang panjang sehingga memakan waktu yang lama (Mahmood, 2012) (Faticihah, 2012). Selain itu, hasil diagnosis cenderung subjektif tergantung pengalaman dokter (Faticihah, 2012). Penghitungan menggunakan mesin memiliki akurasi yang lebih baik daripada penghitungan secara manual. Namun, sel darah merah seringkali dijumpai bertumpuk (*overlapping*) sehingga sulit untuk dianalisis. Masalah ini menyebabkan ketidakakuratan kedua metode tersebut. Masalah sel bertumpuk harus diatasi untuk meningkatkan akurasi penghitungan sel darah merah.

Terdapat beberapa penelitian yang mengembangkan metode identifikasi dan pemisahan sel bertumpuk. (Cheng & Rajapakse, 2009) dan (Yang dkk., 2006) mengusulkan metode *marker-controlled watershed* untuk memisahkan sel bertumpuk pada sel-sel saraf dan *drosophila*, sedangkan (Malpica dkk., 1997) mengaplikasikan metode *watershed* ini pada inti sel yang telah dikelompokkan. Metode ini akan bekerja dengan baik jika *shape-marker* bisa mewakili objek dengan baik. Pemisahan sel bertumpuk pada sel serviks diperkenalkan oleh (Jung dkk., 2010) yang dirumuskan sebagai masalah analisis *cluster* (pengelompokan). Penelitian tersebut melakukan pengelompokan menggunakan *unsupervised Bayesian classifier*. Hasil uji coba menunjukkan bahwa metode tersebut mampu melakukan segmentasi sel bertumpuk lebih baik dibandingkan dengan metode *marker-controlled watershed*. (Maitra dkk., 2012) mengidentifikasi sel darah merah bertumpuk dan tidak bertumpuk, dan memperkirakan jumlahnya menggunakan teknik transformasi Hough. Teknik tersebut mendeteksi sel darah merah dengan mengasumsikan sel sebagai sebuah lingkaran. Pendekatan pada karakteristik geometri sel darah merah diusulkan (Effendy dkk., 2013) untuk mendeteksi sel darah merah bertumpuk. Ide dasar penelitian ini bahwa sel darah merah bertumpuk memiliki karakteristik geometri, luasan dan kelonjongan (eksentrisitas), yang spesifik dibandingkan sel-sel yang tidak bertumpuk. Penggunaan fitur geometri ini dapat mengidentifikasi dan memperkirakan jumlah sel darah merah bertumpuk dan tidak bertumpuk lebih akurat daripada transformasi Hough, namun adanya sel darah merah bertumpuk berderajat tinggi (*high overlapping*) menyebabkan ketidakakuratan penghitungan.

Pada penelitian ini diusulkan identifikasi sel darah merah bertumpuk dengan pendekatan teknik *data mining*. Permasalahan dimodelkan sebagai masalah klasifikasi. Fitur geometri dan warna digunakan. Ide dasar penggunaan fitur warna yaitu terdapat perbedaan karakteristik warna yang signifikan antara sel yang bertumpuk dengan sel tidak bertumpuk. Sel darah merah bertumpuk memiliki ukuran yang lebih besar, eksentrisitas lebih besar, dan warna yang lebih gelap daripada sel tidak bertumpuk. Sedangkan, sel darah merah bertumpuk berderajat tinggi memiliki identitas yang hampir sama dengan sel tidak bertumpuk namun warnanya yang lebih gelap. Berdasarkan hipotesis tersebut, pendekatan *fuzzy* digunakan untuk memahami penggunaan variabel linguistik misalnya ukuran sel besar, eksentrisitas sedang, dan warna sangat gelap. Selanjutnya algoritma *classifier* diaplikasikan untuk mengklasifikasi sel darah merah, sebagai sel tidak bertumpuk, bertumpuk, atau bertumpuk derajat tinggi, berdasarkan fitur yang telah diekstraksi yang ditunjukkan pada Gambar 1.



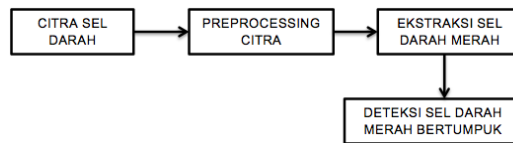
Gambar 1. (a) Sel darah merah tidak bertumpuk, (b) Sel darah merah bertumpuk, (c) Sel darah merah bertumpuk berderajat tinggi

Pohon keputusan *fuzzy* atau *Fuzzy Decision Tree* (FDT) merupakan perluasan algoritma *classifier decision tree* yang memiliki tingkat komprehensif tinggi terhadap perilaku pada sistem *fuzzy* (Chengming, 2007). Ada beberapa metode FDT yang telah dikembangkan seperti FDT berbasis *entropy* (Chengming, 2007) (Umamo dkk., 1994) (Pedrycz, 2005) (Qin dan Lawry, 2005) (Shyi-ming & Fu-Ming, 2008) dan FDT berbasis *gini index* (Chandra & Varghese, 2009).

Metode *classifier* yang akan digunakan pada penelitian ini yaitu FDT berbasis *gini index* (Chandra & Varghese, 2009) dengan modifikasi untuk kasus pencabangan *multi split*. *Gini index* dipilih sebagai pengukuran atribut *split* karena algoritma *entropy* memiliki operasi matematis yang lebih kompleks dibandingkan metode berbasis *gini index*.

2. Metodologi

Proses identifikasi sel darah merah bertumpuk dibagi menjadi 3 proses. Proses yang pertama adalah *preprocessing* citra sel darah. Pada *preprocessing* citra dilakukan transformasi citra masukan hasil akuisisi menjadi citra biner dan dilakukan pula perbaikan kualitas piksel citra. Proses selanjutnya dilakukan pemisahan sel darah merah dengan objek selain sel darah merah sehingga informasi ukuran, eksentrisitas dan warna sel dapat diekstrak. Informasi tersebut digunakan untuk mendeteksi sel bertumpuk. Gambar 2 menunjukkan diagram tahapan-tahapan tersebut.

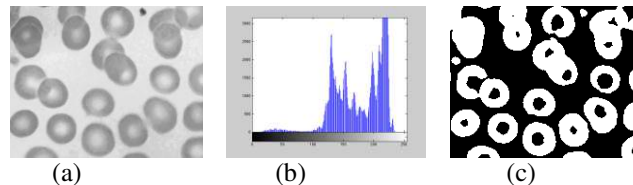


Gambar 2. Diagram tahapan metode yang diusulkan

2.1. Preprocessing citra

Preprocessing bertujuan untuk memperbaiki citra masukan sebelum diolah lebih lanjut. Proses tersebut meliputi filter *noise*, mengubah citra masukan RGB menjadi citra biner berdasarkan nilai *threshold*-nya, dan operasi morfologi yaitu operasi *flood-fill*.

Citra masukan RGB yang berisi *noise* harus dilakukan filter untuk memperhalus citra. Investigasi histogram intensitas warna citra skala keabuan dilakukan. Gambar 3, menunjukkan histogram dengan distribusi bimodal. Puncak yang lebih rendah merepresentasikan piksel latar belakang dan puncak yang lebih tinggi merepresentasikan objek. Nilai optimum (*threshold* τ) yang memisahkan dua puncak perlu diidentifikasi untuk memisahkan objek dan latar belakang. Algoritma *Otsu* (1979) diaplikasikan untuk mendapatkan nilai *threshold* τ .



Gambar 3. (a) Citra skala keabuan, (b) Histogram citra skala keabuan, (c) Citra binernya

Ide dasar dari pendekatan *Otsu* adalah mencari nilai *threshold* yang memaksimalkan *between-class variance* $\sigma_B^2(t)$ (atau meminimumkan *within-class variance*), yang dinyatakan:

$$\sigma_B^2(t) = \frac{[m_T P(t) - m(t)]^2}{P(t)[1 - P(t)]} \quad (1)$$

dengan m_T adalah intensitas rata-rata I_{gray} , I_{gray} adalah citra keabuan sel darah, $m(t)$ adalah rata-rata kumulatif (*cumulative mean*) hingga *gray-level* t , sedangkan $P(t)$ adalah jumlah kumulatif (*cumulative sum*) probabilitas piksel hingga *gray-level* t , masing-masing dinyatakan sebagai:

$$m_T = \sum_{t=0}^{L-1} t p(t) \quad (2)$$

$$m(t) = \sum_{t=0}^{t-1} t p(t) \quad (3)$$

$$P(t) = \sum_{t=0}^{t-1} p(t) \quad (4)$$

$p(t)$ adalah probabilitas *gray-level* t citra I_{gray} yang dirumuskan:

$$p(t) = \frac{n_t}{N}, p(t) \geq 0, \sum_{t=0}^{L-1} p(t) = 1 \quad (5)$$

n_t menyatakan jumlah piksel *gray-level* t citra I_{gray} , L menyatakan jumlah *gray-level* (255), dan N menyatakan total piksel citra I_{gray} . Nilai *threshold* τ adalah nilai t yang memaksimumkan $\sigma_B^2(t)$ sehingga dapat dinyatakan:

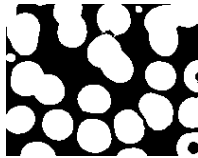
$$\tau = \arg \max_{0 \leq t \leq L-1} \sigma_B^2(t) \quad (6)$$

Citra biner diperoleh dengan mengacu pada nilai *threshold* τ berdasarkan persamaan (7)

$$I_{biner} = \begin{cases} 1, & \text{jika } I_{gray} > \tau \\ 0, & \text{jika tidak} \end{cases} \quad (7)$$

dengan I_{biner} adalah citra biner sel darah.

Pada citra biner tersebut kadang terdapat lubang yang disebabkan iluminasi citra saat proses akuisisi citra. Lubang tersebut akibatnya dianggap sebagai sebuah latar belakang. Untuk mengatasinya digunakan operasi morfologi *flood-fill* yang ditunjukkan pada Gambar 4.



Gambar 4. Citra biner sel darah setelah operasi morfologi *flood-fill*

2.2. Ekstraksi citra sel darah merah

Hampir semua citra sel darah merah mengandung objek yang tidak dibutuhkan yaitu sel darah putih dan platelet. Oleh karena itu, objek yang tidak dibutuhkan tersebut harus dihilangkan. Citra sel darah putih mudah dikenali karena memiliki ukuran yang lebih besar daripada sel darah merah, sedangkan citra platelet memiliki ukuran yang paling kecil sehingga menghilangkannya cukup dengan operasi morfologi *opening*.

Untuk mengekstrak citra sel darah putih digunakan persamaan (8), citra biner sel darah putih diperoleh menggunakan persamaan (9),

$$I_{WBC} = I_B - \frac{1}{2}I_R - \frac{1}{2}I_G \quad (8)$$

$$I_{WBC \text{ biner}} = \begin{cases} 1, & \text{jika } I_{WBC} > 29 \\ 0, & \text{jika tidak} \end{cases} \quad (9)$$

dengan I_{WBC} adalah citra RGB sel darah putih, $I_{WBC \text{ biner}}$ adalah citra biner sel darah putih, I_B adalah komponen biru (*blue*) dari citra masukan RGB, I_R adalah komponen merah (*red*) dari citra masukan RGB, dan I_G adalah komponen hijau (*green*) dari citra masukan RGB, sehingga citra biner sel darah merah dapat diekstrak dengan persamaan (10), dengan $I_{RBC \text{ biner}}$ adalah citra biner sel darah merah.

$$I_{RBC \text{ biner}} = I_{biner} - I_{WBC \text{ biner}} \quad (10)$$

2.3. Deteksi sel darah merah bertumpuk

2.3.1. Ekstraksi fitur sel darah merah

Setelah sel darah merah dipisahkan dari objek yang lain, maka fitur geometri dan warna dapat diekstrak. Luasan didefinisikan sebagai jumlah piksel suatu *region* pada citra. Eksentrisitas merupakan ukuran kelengkungan sebuah elips yang bernilai antara 0 dan 1. Eksentrisitas dirumuskan sebagai rasio antara foci (titik fokus elips) dengan panjang *major axis*.

$$e = 2 \times \frac{\sqrt{\left(\frac{a}{2}\right)^2 - \left(\frac{b}{2}\right)^2}}{a} \quad (11)$$

e adalah nilai eksentrisitas, a adalah panjang *major axis*, dan b adalah panjang *minor axis*.

Sedangkan fitur warna merepresentasikan rata-rata atau *mean* warna sel darah merah pada suatu *region*. Tabel 1 mendeskripsikan fitur-fitur yang digunakan.

Tabel 1. Fitur Sel Darah Merah

| Kategori Fitur | Deskripsi Fitur |
|----------------|----------------------------------|
| Luasan | Luas sel (<i>area</i>), |
| Eksentrisitas | Nilai kelonjongan sel, |
| Warna | Rataan warna sel (<i>mean</i>) |

2.3.2. Klasifikasi sel darah merah bertumpuk menggunakan pohon keputusan fuzzy

2.3.2.1. Dataset

Data citra yang digunakan pada penelitian ini diambil dari basis data *online* citra sel darah perifer yang dimiliki oleh Laboratorium Patologi Internet *Mercer University School of Medicine* (<http://library.med.utah.edu/WebPath/HEMEHTML/HEMEIDX.html>), citra medis internasional CDC (<http://phil.cdc.gov/>), dan *American Society of Hematology* (<http://imagebank.hematology.org>). Citra sel darah tersebut merupakan citra RGB yang memiliki resolusi 504x198 hingga 919x691 piksel.

Sebanyak 15 citra sel darah perifer yang mengandung 750 sel darah merah digunakan sebagai citra data latih. Setelah citra data latih dilakukan *preprocessing* citra, dan ekstraksi sel darah merah (pemisahan sel darah merah dari objek selain sel darah merah), maka fitur (luasan, eksentrisitas, dan warna) dari *region* sel darah merah dapat diekstrak.

Sel darah merah digolongkan ke dalam tiga kelas yaitu kelas 1 (sel darah merah tidak bertumpuk), kelas 2 (sel darah merah bertumpuk), dan kelas 3 (sel darah merah bertumpuk berderajat tinggi). Tabel 2 menunjukkan cuplikan fitur sel darah merah yang telah diekstraksi beserta kelasnya.

Tabel 2. Cuplikan Fitur Sel Darah Merah yang Telah Diekstrak Beserta Kelasnya

| No. Region Sel Darah Merah | Luasan | Eksentrisitas | Warna | Kelas |
|----------------------------|--------|---------------|----------|-------|
| 1 | 587 | 0.6811 | 133.2896 | 1 |
| 2 | 597 | 0.6019 | 131.2831 | 3 |
| 3 | 868 | 0.9206 | 128.4355 | 3 |
| 4 | 895 | 0.5598 | 141.4804 | 1 |
| 5 | 961 | 0.6053 | 141.2268 | 1 |
| 6 | 996 | 0.7978 | 124.5321 | 3 |
| 7 | 1003 | 0.7073 | 184.2393 | 1 |
| 8 | 1109 | 0.6857 | 178.3652 | 1 |
| 9 | 1162 | 0.6625 | 119.9449 | 3 |
| 10 | 1180 | 0.4420 | 159.8627 | 1 |
| 11 | 1244 | 0.4887 | 167.8754 | 1 |
| 12 | 1270 | 0.2199 | 171.9071 | 1 |
| 13 | 1404 | 0.7223 | 193.7486 | 2 |
| 14 | 1642 | 0.8516 | 121.8934 | 3 |
| 15 | 2485 | 0.8537 | 185.8270 | 2 |
| 16 | 2527 | 0.6133 | 141.8556 | 3 |
| 17 | 2683 | 0.8207 | 143.8300 | 2 |
| 18 | 3005 | 0.8272 | 153.2236 | 2 |
| 19 | 3328 | 0.8588 | 197.2849 | 2 |
| 20 | 3894 | 0.8908 | 150.4450 | 2 |

2.3.2.2. Induksi pohon keputusan fuzzy

Setelah informasi fitur data latih telah diekstrak dan telah diberi label kelas, maka tahap selanjutnya adalah membangkitkan aturan klasifikasi menggunakan algoritma *classifier* pohon keputusan fuzzy berbasis *gini index* (Chandra & Varghese, 2009). Pada algoritma pohon keputusan fuzzy berbasis *gini index* ini pra-fuzzifikasi tidak dilakukan, proses fuzzifikasi dilakukan bersamaan dengan induksi pohon keputusan.

2.3.2.2.1. Fuzzifikasi

Proses fuzzifikasi diawali dengan menentukan batas keputusan fuzzy masing-masing atribut. Batas keputusan fuzzy ini berfungsi memisahkan nilai rekod yang kurang dari titik *split* dan lebih dari titik *split*. Dengan kata lain, batas keputusan fuzzy adalah titik *split*. Pemilihan titik *split* didasarkan pada perubahan informasi kelas pada data rekod sebuah atribut yang telah

diurutkan *ascending* sebelumnya (Chandra & Varghese, 2009). Penentuan titik *split* ini merupakan dasar pembentukan fungsi keanggotaan *fuzzy*. Sebagai contoh, titik *split* atribut Luasan dari Tabel 2 ditunjukkan pada Tabel 3.

Tabel 3. Cuplikan Data Fitur Luasan yang Telah Diurutkan *Ascending*

| No. Region Sel Darah Merah | Luasan | Kelas | titik split |
|----------------------------|--------|-------|-------------|
| 1 | 587 | 1 | → 592 |
| 2 | 597 | 3 | |
| 3 | 868 | 3 | → 881.5 |
| 4 | 895 | 1 | |
| 5 | 961 | 1 | → 978.5 |
| 6 | 996 | 3 | → 999.5 |
| 7 | 1003 | 1 | |
| 8 | 1109 | 1 | → 1135.5 |
| 9 | 1162 | 3 | → 1752 |
| 10 | 1180 | 1 | |
| 11 | 1244 | 1 | |
| 12 | 1270 | 1 | → 1337 |
| 13 | 1404 | 2 | → 1523 |
| 14 | 1642 | 3 | → 2063.5 |
| 15 | 2485 | 2 | → 2506 |
| 16 | 2527 | 3 | → 2605 |
| 17 | 2683 | 2 | |
| 18 | 3005 | 2 | |
| 19 | 3328 | 2 | |
| 20 | 3894 | 2 | |

Fungsi keanggotaan *fuzzy* dari fitur luasan menggunakan fungsi keanggotaan *2-term* (Gambar 5). Rekod yang bernilai kurang dari titik *split* p dapat dikatakan berkeanggotaan “kecil” atau “rendah” sedangkan rekod yang bernilai lebih besar dari titik *split* p dapat dikatakan berkeanggotaan “besar” atau “tinggi”. Fungsi keanggotaan jenis ini merepresentasikan *binary split* pada pohon keputusan.

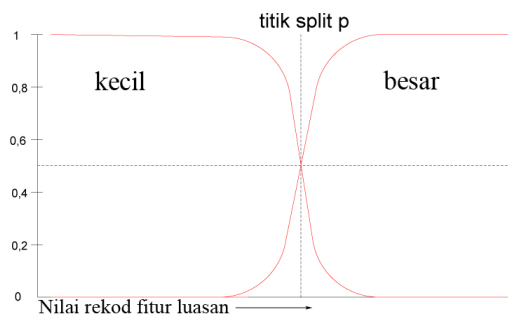
Nilai keanggotaan atribut didapatkan dengan persamaan (12) dan (13) (Chandra & Varghese, 2009),

$$\mu_{Kecil} = 1 - \frac{1}{1 + \exp^{-(\sigma)*(val-p)}} \tag{12}$$

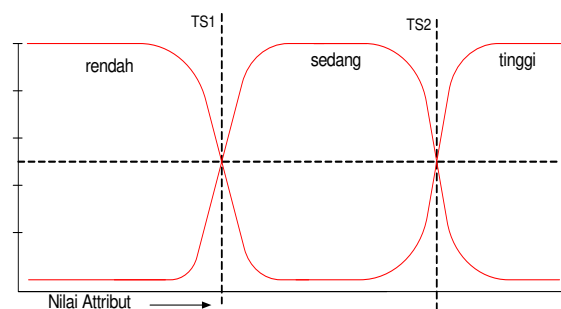
$$\mu_{Besar} = \frac{1}{1 + \exp^{-(\sigma)*(val-p)}} \tag{13}$$

dengan val merupakan nilai rekod pada suatu atribut, σ adalah deviasi standarnya (*standard deviation*), sedangkan μ_{Kecil} adalah nilai keanggotaan untuk nilai rekod yang kurang dari titik *split* p dan μ_{Besar} untuk nilai rekod yang lebih besar dari titik *split* p .

Berbeda dengan luasan, fitur eksentrisitas dan warna menggunakan fungsi keanggotaan *fuzzy* dengan tiga *term* (rendah, sedang, tinggi) seperti pada Gambar 6.



Gambar 5. Fungsi keanggotaan *fuzzy* 2-term dengan titik *split* p



Gambar 6. Fungsi keanggotaan *fuzzy* 3-term dengan dua titik *split*

Ada dua titik *split* yang akan digunakan untuk membangkitkan fungsi keanggotaan fuzzy yaitu titik *split* pertama (*TS1*) dan titik *split* kedua (*TS2*). Sebagai contoh, titik *split* atribut Warna ditunjukkan Tabel 4. Misalkan nilai 132.3 dipilih sebagai titik *split* pertama (*TS1*), maka titik *split* kedua (*TS2*) adalah titik *split* lainnya (selain titik *split* *TS1*), misalnya 142.8.

Tabel 4. Cuplikan Data Fitur Warna yang Telah Diurutkan Ascending

| No. Region Sel Darah Merah | Warna | Kelas | |
|----------------------------|----------|-------|---------------------|
| 9 | 119.9449 | 3 | |
| 14 | 121.8934 | 3 | |
| 6 | 124.5321 | 3 | |
| 3 | 128.4355 | 3 | |
| 2 | 131.2831 | 3 | |
| 1 | 133.2896 | 1 | titik split → 132.3 |
| 5 | 141.2268 | 1 | |
| 4 | 141.4804 | 1 | |
| 16 | 141.8556 | 3 | → 141.7 |
| 17 | 143.83 | 2 | → 142.8 |
| 20 | 150.445 | 2 | |
| 18 | 153.2236 | 2 | |
| 10 | 159.8627 | 1 | → 156.5 |
| 11 | 167.8754 | 1 | |
| 12 | 171.9071 | 1 | |
| 8 | 178.3652 | 1 | |
| 7 | 184.2393 | 1 | |
| 15 | 185.827 | 2 | → 185 |
| 13 | 193.7486 | 2 | |
| 19 | 197.2849 | 2 | |

Fungsi keanggotaan fuzzy dengan *term* ini dapat dibangkitkan dengan memodifikasi persamaan (12) dan (13) menjadi

$$\mu_{Rendah} = 1 - \frac{1}{1 + \exp^{(-(\sigma)*(val-TS1))}} \quad (14)$$

$$\mu_{Sedang} = \left(1 - \frac{1}{1 + \exp^{(-(\sigma)*(val-TS2))}}\right) \times \left(\frac{1}{1 + \exp^{(-(\sigma)*(val-TS1))}}\right) \quad (15)$$

$$\mu_{Tinggi} = \frac{1}{1 + \exp^{(-(\sigma)*(val-TS2))}} \quad (16)$$

dengan μ_{Rendah} adalah nilai keanggotaan untuk nilai rekod yang kurang dari titik *split* *TS1*, μ_{Sedang} untuk nilai rekod antara *TS1* dan *TS2*, μ_{Sedang} untuk nilai rekod antara *TS1* dan *TS2*, dan μ_{Tinggi} untuk nilai rekod yang lebih besar dari titik *split* *TS2*. Fungsi keanggotaan jenis ini merepresentasikan *multisplit* pada pohon keputusan.

Lebih singkatnya, algoritma fuzzifikasi dirumuskan sebagai berikut:

fungsi fuzzifikasi 2-term (atribut *a*, titik *split* *v*) (Chandra & Varghese, 2009)

1. Hitung deviasi standar σ atribut *a*
2. Untuk setiap nilai rekod *val* pada atribut *a*,
3. Hitung nilai keanggotaan fuzzy atribut menggunakan persamaan (12) untuk $val \leq v$ dan persamaan (13) untuk $val > v$
4. selesai
5. selesai

fungsi fuzzifikasi 3-term (atribut *a*, titik *split* *v1*, titik *split* *v2*)

1. Hitung deviasi standar σ atribut *a*
2. Untuk setiap nilai rekod *val* pada atribut *a*,
3. Hitung nilai keanggotaan fuzzy atribut menggunakan persamaan (14) untuk $val \leq v1$, persamaan (15) untuk $v1 < val < v2$, dan persamaan (16) untuk $val \geq v2$
4. selesai
5. selesai

2.3.2.2.1. Pembangkitan pohon keputusan *fuzzy*

Metode induksi pohon keputusan *fuzzy* ini merupakan perluasan algoritma pohon keputusan SLIQ yang menggunakan *gini index* untuk mengevaluasi *split* terbaik. Jika pada algoritma SLIQ nilai *gini index* dihitung dari jumlah rekod pada suatu atribut yang berasosiasi dengan kelas tertentu, maka pada pohon keputusan *fuzzy* berbasis *gini*, *gini index* dihitung dari nilai keanggotaan *fuzzy* dari rekod pada suatu atribut yang berasosiasi dengan kelas tertentu.

Algoritma konstruksi pohon keputusan *fuzzy* diberikan sebagai berikut (Chandra & Varghese, 2009) dengan modifikasi,

- 1 Jika L merupakan *root* maka
- 2 Nilai keanggotaan fuzzy tiap rekod data diberi nilai $1/\text{jumlah kelas}$.
- 3 Jika tidak,
- 4 Nilai keanggotaan fuzzy tiap rekod data diperbarui dengan mengalikannya dengan nilai keanggotaan fuzzy atribut yang dipilih.
- 5 Selesai
- 6 Untuk setiap node di level L , lakukan
- 7 Jika kriteria split tidak tercapai, maka
- 8 Untuk setiap atribut a pada dataset lakukan
- 9 Urutkan nilai atribut rekod
- 10 Untuk setiap nilai titik split v lakukan
- 11 **Evaluasi Gini Index** (a, v) untuk binary split
- Evaluasi Gini Index** ($a, v1, v2$) untuk multisplit
- 12 Selesai
- 13 Selesai
- 14 Pilih atribut a dengan titik split v yang memiliki nilai Gini Index terkecil sebagai atribut split
- 15 Jadikan atribut terpilih sebagai *leaf*
- 16 Ulangi konstruksi pohon keputusan untuk level $L+1$
- 17 Selesai
- 18 Selesai

Fungsi Evaluasi *Gini Index* (atribut a , titik split v) untuk binary split

- 1 Untuk setiap nilai rekod dari atribut a lakukan
- 2 fuzzifikasi 2-term (a, v)
- 3 Selesai
- 4 Evaluasi Gini Index menggunakan nilai keanggotaan *fuzzy* atribut

Fungsi Evaluasi *Gini Index* (atribut a , titik split $v1$, titik split $v2$) untuk multisplit

- 1 Untuk setiap nilai rekod dari atribut a
- 2 fuzzifikasi 3-term ($a, v1, v2$)
- 3 Selesai
- 4 Evaluasi Gini Index menggunakan nilai keanggotaan *fuzzy* atribut

Masing-masing rekod pada *dataset* dihitung nilai keanggotaan *fuzzy*-nya. Pada keadaan awal (inisial), nilai keanggotaan *fuzzy* masing-masing rekod diberi nilai = $1/\text{jumlah kelas}$. Untuk node lainnya (selain *root*) nilai keanggotaan *fuzzy* diperbarui bergantung pada atribut yang dipilih sebagai atribut *split*. Setiap nilai atribut diurutkan terlebih dulu. Nilai titik *split* dihitung ketika informasi kelas berubah. Nilai keanggotaan atribut pada tiap rekod dihitung menggunakan fungsi fuzzifikasi.

Selanjutnya, *gini index* dievaluasi untuk tiap nilai titik *split* untuk semua atribut. Perhitungan *gini index* dirumuskan sebagai berikut,

$$D(x_j) = \sum_{v=1}^V \frac{N^{(v)}}{N^{(u)}} \left(1 - \sum_{k=1}^C \frac{N^{(v)_k}}{N^{(v)}} \right)^2 \quad (17)$$

dengan C adalah jumlah kelas; V adalah jumlah partisi; $N^{(u)}$ adalah jumlah nilai keanggotaan *fuzzy* rekod pada *dataset* jika x_j dipilih sebagai titik *split*; $N^{(v)}$ adalah jumlah nilai keanggotaan

fuzzy pada partisi ke- v ; $N_k^{(v)}$ adalah jumlah perkalian nilai keanggotaan atribut dan nilai keanggotaan dari rekod yang berhubungan untuk kelas ω_k pada partisi ke- v .

Sebagai contoh, Tabel 5 menunjukkan penghitungan *gini index* pada atribut Warna dengan titik *split* pertama (TS1) 132.3 dan titik *split* kedua (TS2) 141.7. Nilai $N^{(u)}$ adalah total nilai dari kolom 5 yaitu 6.6604. Nilai $N^{(1)}$ adalah total nilai dari kolom Y1 yaitu 1.667. Nilai $N^{(2)}$ adalah total nilai dari kolom Y2 yaitu 0.9999. Nilai $N^{(3)}$ adalah total nilai dari kolom Z yaitu 4.0000. Nilai $N_{\omega_1}^{(1)}$ adalah nilai $N^{(1)}$ yang merupakan anggota kelas 1 yaitu 0 (mendekati 0 karena terdapat pembulatan dalam tabel). Nilai $N_{\omega_2}^{(1)}$ adalah nilai $N^{(1)}$ yang merupakan anggota kelas 2 yaitu 0. Nilai $N_{\omega_3}^{(1)}$ adalah nilai $N^{(1)}$ yang merupakan anggota kelas 3 yaitu 1.6667. Demikian seterusnya, sehingga nilai *gini index*nya adalah 0.3423.

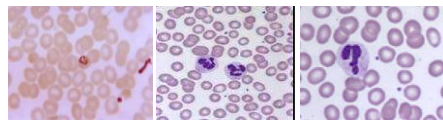
Atribut *split* terbaik, yaitu atribut dengan *gini index* terkecil, dipilih sebagai *split node*. Nilai keanggotaan fuzzy rekod kemudian diperbarui menggunakan nilai keanggotaan fuzzy atribut dari atribut terpilih.

2.3.2.2.3. Aturan yang dibangkitkan

Aturan yang telah dibangkitkan berdasarkan data latih menggunakan pohon keputusan fuzzy berbasis *gini index* ditunjukkan Gambar 8. Aturan berupa pernyataan *IF..THEN*. Misalkan salah satu aturan yang dapat diturunkan berdasarkan Gambar 8 yaitu *IF Luasan > 1135 AND Eksentrisitas > 0.71 THEN sel darah merah bertumpuk*.

3. Hasil dan Pembahasan

Metode yang diusulkan ini telah diuji dengan 10 citra mikroskopik sel darah merah yang mengandung 532 sel darah merah dengan variasi kelas yang berbeda. Untuk mengetahui akurasi dilakukan penghitungan secara manual terhadap seluruh data uji yang hasilnya dibandingkan dengan hasil penghitungan metode yang diusulkan, metode pohon keputusan fuzzy binary split (Chandra & Varghese, 2009), ditunjukkan oleh Gambar 9, dan penghitungan sel darah merah menggunakan fitur geometri (Effendy dkk., 2013).



Gambar 7. Contoh citra uji

Sel darah merah yang telah diklasifikasi ke dalam sel tidak bertumpuk, sel bertumpuk dan sel bertumpuk berderajat tinggi kemudian dihitung jumlahnya. Estimasi jumlah sel darah pda sel darah merah bertumpuk menggunakan persamaan yang diusulkan oleh Effendy (2013),

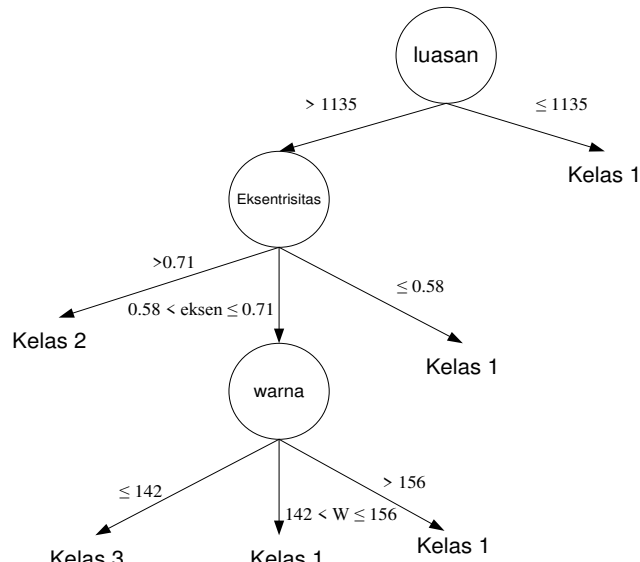
$$\text{Jumlah sel darah} = \frac{A}{B - (0.5 \times \text{std}(B))} \quad (18)$$

dengan A adalah ukuran (luasan) sel bertumpuk, B adalah rata-rata luasan sel tidak bertumpuk, dan std adalah deviasi standar. Sedangkan, sel bertumpuk berderajat tinggi memiliki jumlah sel darah merah sebanyak dua sel.

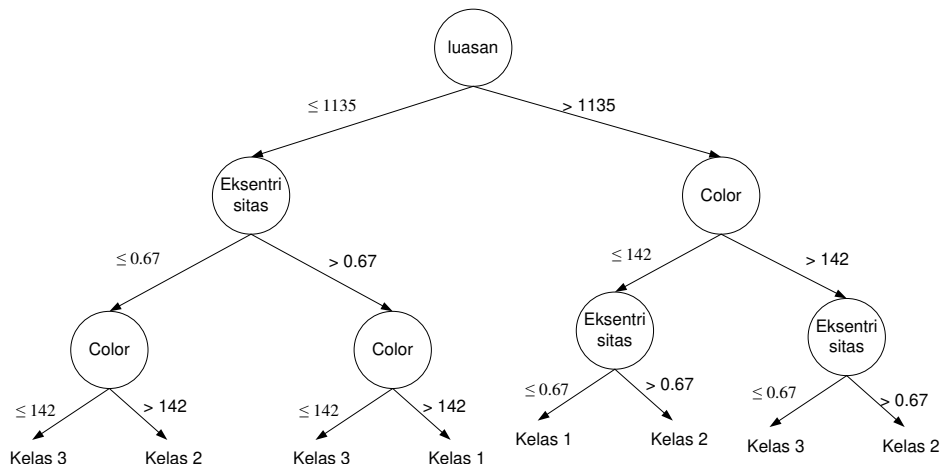
Akurasi metode yang diusulkan didasarkan pada presentaseperbandingan hasil penghitungan menggunakan metode yang diusulkan (pohon keputusan fuzzy dengan pencabangan *multi split*) dengan hasil penghitungan manual. Hasil perbandingan penghitungan dengan metode lain dan akurasi metode yang diusulkan ditunjukkan pada Tabel 5 dan Tabel 6.

Pohon keputusan yang dibangun dengan pencabangan *binary split* seperti pada Gambar 9 mempunyai cabang yang banyak jika dibandingkan dengan pohon keputusan yang dibangun oleh metode yang diusulkan (dengan pencabangan *multi split*) seperti pada Gambar 8. Akurasi penghitungan menggunakan pohon keputusan fuzzy binary split lebih rendah dibanding dengan menggunakan pencabangan *multi split*. Hal ini dapat disebabkan fitur eksentrisitas hanya mengakomodir *2-term*, berbeda dengan pencabangan *multi split* yang menggunakan *3-term*

sehingga dapat terjadi *over counting*. Contohnya, sebuah *region* citra sel darah merah yang memiliki nilai luasan 895, nilai eksentrisitas 0.56 dan nilai rata-ran warna 141.5 akan diklasifikasikan sebagai sel darah merah bertumpuk berderajat tinggi sehingga dihitung jumlah sel darah merah adalah dua. Bila menggunakan pencabangan *multi split* maka diklasifikasikan sebagai sel darah merah tak bertumpuk (jumlah sel satu). Pengamatan manual mengatakan bahwa *region* sel darah merah tersebut tak bertumpuk.



Gambar 8. Pohon keputusan metode yang diusulkan



Gambar 9. Pohon keputusan fuzzy binary split

Peningkatan akurasi penghitungan juga terjadi bila dibandingkan dengan hanya menggunakan fitur geometri (luasan dan eksentrisitas). Hal ini terjadi bila pada citra sel darah terdapat sel darah merah bertumpuk berderajat tinggi. Sel darah merah bertumpuk berderajat tinggi memiliki ukuran (luasan) dan eksentrisitas yang hampir sama dengan sel tak bertumpuk. Untuk mendeteksi dengan fitur geometri saja tidak cukup serta estimasi jumlah sel darah merah juga tidak dapat menggunakan persamaan (12), karena jumlah sel darah merah akan diestimasi jumlahnya satu. Melalui pengamatan, sel darah merah bertumpuk berderajat tinggi memiliki luasan dan eksentrisitas seperti sel tak bertumpuk namun warnanya lebih gelap dan jumlah sel yang bertumpuk ada dua sel.

Akurasi yang terbaik pada uji coba dengan menggunakan metode yang diusulkan adalah mencapai 100% sedangkan yang terendah adalah 84.85% dan rata-rata akurasi untuk keseluruhan data uji coba 96.14%.

Tabel 5. Perbandingan Hasil Penghitungan

| Citra | Perhitungan | | | |
|-------|-------------|------------------------------------|----------------|-----------------------|
| | Manual | Pohon Keputusan Fuzzy binary split | Fitur Geometri | Metode yang Diusulkan |
| 1 | 89 | 95 | 89 | 89 |
| 2 | 34 | 37 | 33 | 33 |
| 3 | 36 | 35 | 35 | 35 |
| 4 | 77 | 81 | 73 | 74 |
| 5 | 80 | 82 | 79 | 79 |
| 6 | 35 | 36 | 32 | 33 |
| 7 | 74 | 77 | 71 | 73 |
| 8 | 41 | 43 | 39 | 40 |
| 9 | 33 | 34 | 28 | 28 |
| 10 | 33 | 35 | 32 | 32 |

Tabel 6. Akurasi Hasil Penghitungan

| Citra | Perhitungan | | |
|-------|-------------|-----------------------|--|
| | Manual | Metode yang Diusulkan | Akurasi Penghitungan menggunakan Metode yang Diusulkan |
| 1 | 89 | 89 | 100% |
| 2 | 34 | 33 | 97.05% |
| 3 | 36 | 35 | 97.22% |
| 4 | 77 | 74 | 96.10% |
| 5 | 80 | 79 | 98.75% |
| 6 | 35 | 33 | 94.28% |
| 7 | 74 | 73 | 98.65% |
| 8 | 41 | 40 | 97.56% |
| 9 | 33 | 28 | 84.85% |
| 10 | 33 | 32 | 96.97% |

Tabel 7. Penghitungan Nilai Keanggotaan Fuzzy pada Cuplikan Fitur Warna dengan nilai TS1 132.3 dan nilai TS2 141.7

| No Region Sel Darah Merah | Nilai atribut warna | X | 1 | Y1 | Y2 | Z | 2 | 3 | 4 | 5 |
|---------------------------|---------------------|-----|---|----|---------|---------|---------|---------|---------|---------|
| 9 | 119.9449 | 0.3 | 3 | 1 | 0 | 0 | 0.33333 | 0 | 0 | 0.33333 |
| 14 | 121.8934 | 0.3 | 3 | 1 | 0 | 0 | 0.33333 | 0 | 0 | 0.33333 |
| 6 | 124.5321 | 0.3 | 3 | 1 | 0 | 0 | 0.33333 | 0 | 0 | 0.33333 |
| 3 | 128.4355 | 0.3 | 3 | 1 | 0 | 0 | 0.33333 | 0 | 0 | 0.33333 |
| 2 | 131.2831 | 0.3 | 3 | 1 | 0 | 0 | 0.33333 | 0 | 0 | 0.33333 |
| 1 | 133.2896 | 0.3 | 1 | 0 | 1 | 0 | 0 | 0.33333 | 0 | 0.33333 |
| 5 | 141.2268 | 0.3 | 1 | 0 | 0.99998 | 0.00002 | 0 | 0.33333 | 0.00001 | 0.33333 |
| 4 | 141.4804 | 0.3 | 1 | 0 | 0.99068 | 0.00932 | 0 | 0.33023 | 0.00311 | 0.33023 |
| 16 | 141.8556 | 0.3 | 3 | 0 | 0.00932 | 0.99068 | 0 | 0.00311 | 0.33023 | 0.33023 |
| 17 | 143.83 | 0.3 | 2 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 20 | 150.445 | 0.3 | 2 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 18 | 153.2236 | 0.3 | 2 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 10 | 159.8627 | 0.3 | 1 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 11 | 167.8754 | 0.3 | 1 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 12 | 171.9071 | 0.3 | 1 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 8 | 178.3652 | 0.3 | 1 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 7 | 184.2393 | 0.3 | 1 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 15 | 185.827 | 0.3 | 2 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 13 | 193.7486 | 0.3 | 2 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |
| 19 | 197.2849 | 0.3 | 2 | 0 | 0 | 1 | 0 | 0 | 0.33333 | 0.33333 |

Keterangan:

X adalah nilai keanggotaan fuzzy rekod sebelum split; 1 adalah kelas; Y1 adalah nilai fuzzy atribut Warna untuk nilai rekod kurang dari TS1 (μ_{Rendah}); Y2 adalah nilai fuzzy atribut Warna untuk rekod antara TS1 dan TS2 (μ_{Sedang}); Z adalah nilai fuzzy atribut Warna untuk nilai rekod lebih besar dari TS2 (μ_{Tinggi}); 2 adalah hasil perkalian X dan Y1; 3 adalah hasil perkalian X dan Y2; 4 adalah hasil perkalian X dan Z; 5 adalah nilai keanggotaan fuzzy rekod bila dipilih atribut Warna dengan titik split TS1 dan titik split TS2.

4. Kesimpulan

Pada penelitian ini dilakukan identifikasi sel darah merah bertumpuk menggunakan pohon keputusan fuzzy berbasis gini index. Tiga fitur digunakan yaitu fitur luasan, eksentrisitas dan warna. Fitur yang diekstraksi dari citra data latih sel darah merah dapat digunakan untuk membangkitkan aturan klasifikasi. Aturan klasifikasi dibangkitkan berdasarkan pohon

keputusan *fuzzy* berbasis *gini index*. Pencabangan *multi split* diusulkan pada penelitian ini. Hasil pengujian menunjukkan bila pohon keputusan menggunakan pencabangan *binary split* maka yang akan terjadi adalah *over counting*. Kasus ini dapat terjadi karena adanya misklasifikasi, yaitu sel darah merah tak bertumpuk dapat diklasifikasikan sebagai sel darah bertumpuk berderajat tinggi sehingga terjadi kesalahan hasil penghitungan.

Penggunaan fitur geometri dan warna juga terbukti dapat meningkatkan akurasi penghitungan sebesar 1.04% dibanding dengan fitur geometri saja, disebabkan kehadiran sel darah merah bertumpuk yang tidak dapat dideteksi hanya menggunakan fitur geometri. Diharapkan dengan peningkatan akurasi ini, akurasi diagnosis penyakit yang dipengaruhi oleh jumlah sel darah merah dapat meningkat pula.

Referensi

- Chandra, B. & Varghese, P. 2009. Fuzzifying Gini Index based decision trees. *Expert Systems with Applications*, 36(4): 8549-8559.
- Cheng, J. & Rajapakse, J. 2009. Segmentation of clustered nuclei with shape markers and marking function. *IEEE Trans. Biomed. Eng.*, 56(3): 741 -748.
- Chengming Qi. 2007. *A New Partition Criterion for Fuzzy Decision Tree Algorithm*. Makalah disajikan dalam Intelligent Information Technology Application, Workshop on , Desember.
- Effendy, F., Faticah, C., & Purwitasari, D. 2013. *Identifikasi Sel Bertumpuk Berdasarkan Fitur Geometri pada Sel Darah Merah*. Makalah disajikan dalam Seminar Nasional Pascasarjana XI, Institut Teknologi Sepuluh Nopember, Surabaya, Agustus.
- Faticah, C., Tangel, M. L., Yan, F., Betancourt, J. P., Widyanto, M. R., Dong, F., & Hirota. 2012. K.Fuzzy Feature Representation for White Blood Cell Differential Counting in Acute Leukemia Diagnosis. *Submission to International Journal of Control, Automation, and Systems*.
- Jung, C., Kim, C., Chae, S. W. & Oh, S. 2010. Unsupervised segmentation of overlapped nuclei using Bayesian classification. *IEEE Trans. Biomed. Eng.*, 57(12): 2825 -2832.
- Mahmood, N. H. & Mansor, M. A. 2012. Red Blood Estimation Using Hough Transform Technique. *Signal and Image Processing: An International Journal (SIPIJ)*, 3(2): 53-64.
- Maitra, M., Gupta, R. K., & Mukherjee, M. 2012. Detection and Counting of Red Blood Cells in Blood Cell Images using Hough Transform. *International Journal of Computer Application*, 53(16): 18–22.
- Malpica, N., de Solorzano, C.O., Vaquero, J.J., Santos, A., Vallcorba, I., Garcia-Sagredo, J.M., & del Pozo, F. 1997. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry*, 28(4): 289–297.
- Otsu, N. 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1): 62–66.
- Pedrycz, W. 2005. C-fuzzy decision trees. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 35(4): 498-511.
- Qin, Z. & Lawry, J. 2005. Decision tree learning with fuzzy labels. *Information Sciences*, 172, (1-2): 91-129.
- Shyi-Ming, C. & Fu-Ming, T. 2008. Generating fuzzy rules from training instances for fuzzy classification systems. *Expert Systems with Applications*, 35(3): 611-621.
- Umamo, M., Okamoto, H., Hatono, I., Tamura, H., Kawachi, F., Umedzu, S., & Kinoshita, J. 1994. *Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems*. Makalah disajikan dalam the Third IEEE Conference on Fuzzy Systems, 1994. IEEE World Congress on Computational Intelligence, Juni.
- Yang, X., Li, H., & Zhou, X. 2006. Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy. *IEEE Trans. Circuits Syst.*, 53(11): 2405 -2414.