

IDENTIFIKASI PENYAKIT LEUKIMIA AKUT PADA CITRA DARAH MIKROSKOPIS

Oleh: Indah Susilawati

Staf Pengajar Program Studi Teknik Informatika Universitas Mercu Buana Yogyakarta
Jl. Wates Km. 10 Yogyakarta

Abstrak

Terdapat berbagai jenis leukemia, diantaranya adalah penyakit leukemia limphoblastik akut (ALL: Acute Lymphoblastic Leukemia) yang merupakan jenis leukemia yang paling sering terjadi pada anak-anak. Identifikasi sejak dini gejala ALL pada pasien dapat meningkatkan kemungkinan penyembuhan penyakit jenis ini. Penyakit leukemia dapat diidentifikasi secara otomatis melalui tes tertentu seperti cytogenetics and immunophenotyping dan klasifikasi sel secara morfologis oleh seorang operator yang berpengalaman dengan mengamati citra darah mikroskopis. Cara tersebut mempunyai akurasi yang sangat baik namun kekurangannya terletak pada prosesnya yang lambat dan akurasinya menjadi tidak terstandarisasi karena sangat tergantung pada tingkat keahlian/pengalaman operator dan juga tingkat kelelahan. Pada penelitian ini dilakukan identifikasi penyakit leukemia akut (ALL) menggunakan cara-cara pengolahan citra digital dan classifier kNN. Classifier kNN membedakan sel darah putih normal dan abnormal (ALL) berdasarkan tiga ciri yaitu ciri rerata, varians, dan standar deviasi yang diekstrak dari hasil segmentasi sel darah putih. Hasil penelitian menunjukkan bahwa kinerja classifier kNN mencapai maksimum hingga 94% menggunakan konstanta $k = 5$.

Kata Kunci: *contrast stretching, pengambangan, segmentasi, ALL, kNN.*

1. Pendahuluan

Sebagai cairan tubuh yang sangat penting bagi kehidupan manusia, darah dapat digunakan untuk mengetahui kondisi kesehatan seseorang melalui berbagai tes dan analisis darah. Misalnya analisis terkait penghitungan jumlah sel-sel darah baik sel darah merah, sel darah putih, maupun trombosit. Jumlah sel-sel darah tersebut selanjutnya dapat digunakan sebagai salah satu petunjuk untuk diagnosis penyakit yang diderita oleh pasien.

Perkembangan teknologi mikroskop elektron memungkinkan akuisisi citra darah mikroskopis secara digital untuk selanjutnya dilakukan analisis; dan juga menjadi dimungkinkan untuk dilakukan analisis oleh operator dari jarak jauh (dengan mengirimkan citra darah tanpa harus melihat langsung sampel darah pasien). Di lain pihak, teknik pengolahan citra digital dan teknik klasifikasi serta

pengenalan pola menyediakan berbagai algoritma yang dapat digunakan untuk membantu mengolah dan mengidentifikasi penyakit darah dengan melakukan analisis citra darah.

Salah satu penyakit darah adalah yang terkait dengan penyakit sel darah putih yaitu kanker darah atau leukemia. Terdapat berbagai jenis leukemia, diantaranya adalah penyakit leukemia limphoblastik akut (ALL: Acute Lymphoblastic Leukemia) yang merupakan jenis leukemia yang paling sering terjadi pada anak-anak. Identifikasi sejak dini gejala ALL pada pasien dapat meningkatkan kemungkinan penyembuhan penyakit jenis ini. Penyakit leukemia dapat diidentifikasi secara otomatis melalui tes tertentu seperti *cytogenetics and immunophenotyping* dan klasifikasi sel secara morfologis oleh seorang operator yang berpengalaman dengan mengamati citra darah mikroskopis. Cara tersebut

mempunyai akurasi yang sangat baik namun kekurangannya terletak pada prosesnya yang lambat dan akurasinya menjadi tidak terstandarisasi karena sangat tergantung pada tingkat keahlian/pengalaman operator dan juga tingkat kelelahan.

Pada penelitian ini dilakukan identifikasi penyakit leukemia akut menggunakan cara-cara pengolahan citra digital dan *classifier* kNN. Dengan cara ini maka proses identifikasi penyakit dapat dilakukan dengan lebih cepat dan akurasinya menjadi terstandarisasi karena tidak tergantung pada tingkat keahlian/pengalaman operator dan juga tingkat kelelahannya.

Penelitian sebelumnya yang terkait dengan penelitian ini adalah penelitian yang dilakukan oleh Adnan Khashman (2008) berjudul "*Blood Cell Identification Using A Simple Neural Network*". Dalam penelitian ini dilakukan identifikasi sel darah merah, sel darah putih, dan trombosit. Fitur ciri diekstrak menggunakan *global pattern averaging* dan sebuah jaringan saraf untuk mengklasifikasikan jenis-jenis sel.

Hiremath dkk (2010) dalam penelitiannya berjudul "*Automated Identification and Classification of White Blood Cells (Leucocytes) in Digital Microscopic Images*", melakukan identifikasi dan klasifikasi jenis-jenis sel darah putih menjadi tiga jenis yaitu limposit, monosit, dan netrofil. Identifikasi dan klasifikasi dilakukan menggunakan metode segmentasi berdasar warna dan ekstraksi ciri geometris.

Penelitian yang lain dilakukan oleh Vincenzo Piuri dan Fabio Scotti (2004) dengan judul "*Morphological Classification of Blood Leucocytes by Microscope Images*". Dalam penelitian ini dilakukan deteksi dan klasifikasi jenis-jenis sel darah putih secara otomatis. Jenis-jenis sel darah putih diklasifikasikan menjadi lima jenis yaitu basofil, eosinofil, limposit, monosit, dan netrofil. Sistem yang dibangun mula-

mula memisahkan sel darah putih dari sel-sel darah yang lain, kemudian dilakukan ekstraksi indeks-indeks morfologinya untuk selanjutnya diklasifikasikan menggunakan *neural classifier*.

Lorenzo Putzu dan Cecilia Di Ruberto melakukan penelitian berjudul "*White Blood Cells Identification and Counting from Microscopic Blood Image*" pada tahun 2013. Pada penelitiannya terlebih dahulu dilakukan identifikasi sel-sel darah putih dengan memisahkannya dari jenis sel-sel darah yang lain. Selanjutnya dilakukan proses pemisahan sel-sel darah putih yang saling berhimpitan untuk kemudian dilakukan penghitungan jumlah sel darah putih. Dalam analisisnya peneliti mengkonversi citra darah mikroskopis ke model warna CMYK.

Fabio Scotti (2005) melakukan penelitian berjudul "*Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images*". Sistem yang dibangun mula-mula memisahkan leukosit dari sel-sel darah yang lain, kemudian memilah sel-sel limposit (yang menjadi perhatian untuk identifikasi leukemia akut), mengevaluasi indeks-indeks morfologi dari sel-sel tersebut, dan akhirnya menentukan apakah teridentifikasi leukemia akut atau tidak.

2. Metode Penelitian

2.1. Bahan dan Alat Penelitian

Bahan berupa citra darah mikroskopis yang digunakan diunduh dari basis data dengan alamat <http://www.dti.unimi.it/fscott/all/>. Total digunakan 200 citra darah putih (*leukosit*) terdiri atas 2 kelas yaitu kelas citra yang mengandung sel darah putih normal saja dan kelas citra yang mengandung sel darah putih abnormal (ALL).

2.2. Metodologi

Dalam penelitian ini, langkah-langkah yang dilakukan diperlihatkan pada Gambar 1.



Gambar 1. Blok Diagram Penelitian

Perentangan kontras (*contrast stretching*) dilakukan menggunakan persamaan (1) dan bertujuan supaya citra mempunyai kontras yang lebih baik. Hal ini dilakukan untuk kepentingan kemudahan proses segmentasi.

$$y = \left(\frac{x - a}{b - a} \right) (d - c) + c \quad \dots \dots \dots (1)$$

dengan x adalah piksel citra input, y adalah piksel output, a adalah nilai minimum piksel pada citra input, b adalah nilai maksimum piksel pada citra input, c adalah nilai minimum citra output (ditentukan), dan d adalah nilai maksimum citra output (ditentukan). *Contrast stretching* dilakukan setelah sebelumnya mengubah citra warna menjadi citra aras keabuan.

Untuk kepentingan proses pengambangan, maka hal yang paling penting adalah menentukan nilai ambang yang paling tepat untuk semua citra yang digunakan dalam penelitian, karena dalam penelitian ini hanya digunakan satu nilai ambang yang sama (*non adaptif*). Perkiraaan nilai ambang ditentukan dengan cara membuat histogram dari nilai piksel pada baris ke-128. Dari histogram yang diperoleh maka nilai ambang yang akan digunakan ditentukan dengan menentukan nilai piksel dimana terjadi perubahan yang signifikan pada frekuensi kemunculannya.

Segmentasi bertujuan untuk memisahkan sel darah putih dari latar belakangnya dan dari obyek-obyek lain (sel darah merah dan lain-lain) dengan mempertahankan warna

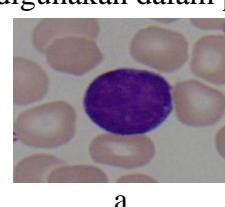
obyek dan membuat latar belakang menjadi hitam (nilai piksel = 0). Segmentasi dilakukan dengan cara yang sederhana yaitu mengalikan citra asli dengan citra hasil pengambangan secara *element wise*.

Tahap ekstraksi ciri berupa proses pengambilan ciri yang akan digunakan sebagai input *classifier* kNN diekstrak dari sel darah putih atau leukosit yang telah disegmentasi pada tahap sebelumnya. Dalam penelitian ini diekstrak tiga ciri yaitu rerata, varians dan standar deviasi; selanjutnya ketiga ciri tersebut dinyatakan dalam bentuk vektor ciri dan digunakan sebagai masukan *classifier* kNN.

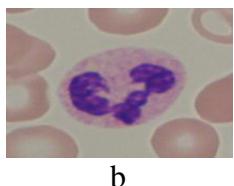
Klasifikasi bertujuan untuk membedakan citra leukosit normal dengan citra leukosit abnormal (ALL). Citra leukosit normal dikelompokkan dalam kelas 1 dan citra leukosit abnormal (ALL) dikelompokkan dalam kelas 2. Algoritma kNN melakukan klasifikasi terhadap obyek berdasarkan data pelatihan yang jaraknya paling dekat dengan objek yang akan diklasifikasikan. Dalam penelitian ini jarak obyek ke tetangga dihitung berdasarkan jarak Euclidean. Pada fase pelatihan, algoritma kNN melakukan penyimpanan vektor-vektor ciri dan klasifikasi dari data pelatihan. Pada fase klasifikasi atau pengujian, dihitung jarak dari vektor ciri obyek yang baru terhadap seluruh vektor data pelatihan, dan diambil sebanyak k jarak yang paling dekat. Obyek yang baru ini diklasifikasikan berdasar jumlah kelas yang terbanyak (dari k jarak terdekat yang diambil). Dalam penelitian ini digunakan nilai k yang bervariasi yaitu 3, 5, 7, 9, dan 11.

3. Hasil dan Pembahasan

Gambar 2. memperlihatkan dua citra darah putih yang digunakan dalam penelitian.



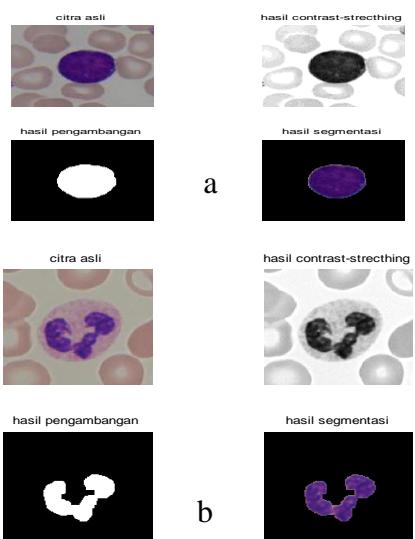
a



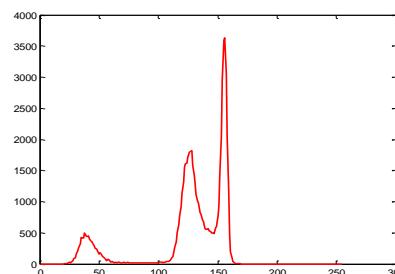
Gambar 2. (a) Citra leukosit abnormal (ALL) Im005_1.tif, (b) Citra leukosit normal Im135_0.tif

Gambar 3. memperlihatkan hasil proses yang dimulai dari perentangan kontras (*contrast stretching*) hingga segmentasi. Perkiraan nilai ambang ditentukan dengan cara mengevaluasi nilai ambang relatif yang paling cocok untuk memisahkan sel darah putih atau leukosit dari bagian-bagian darah yang lain. Gambar 4. memperlihatkan grafik frekuensi kemunculan nilai piksel pada baris ke-128 dari citra Im135_0.tif.

Berdasarkan pengamatan, maka dipilih nilai ambang (*threshold value*) sama dengan 110. Hal ini dilakukan dengan mengamati perubahan nilai piksel pada grafik, dimana perubahan lereng grafik terjadi kira-kira di sekitar nilai piksel 100 – 125. Pengambangan mengubah nilai piksel yang lebih besar atau sama dengan 110 menjadi bernilai 0 dan nilai piksel yang lebih kecil dari 110 menjadi bernilai 1.



Gambar 3. Hasil *contrast stretching* hingga segmentasi (a) citra Im005_1.tif, (b) citra Im135_0.tif



Gambar 4. Grafik frekuensi kemunculan nilai piksel pada citra Im135_0.tif (pada baris ke-128)

Dalam penelitian ini ciri yang digunakan sebagai input *classifier* kNN diekstrak dari sel darah putih atau leukosit yang telah disegmentasi. Diekstrak tiga ciri yaitu rerata, varians dan standar deviasi. Proses selanjutnya adalah vektorisasi dimana ketiga ciri yang telah diekstrak dinyatakan dalam bentuk vektor ciri dan digunakan sebagai masukan *classifier* kNN. Vektor ciri dinyatakan dengan vektor kolom sebagai berikut.

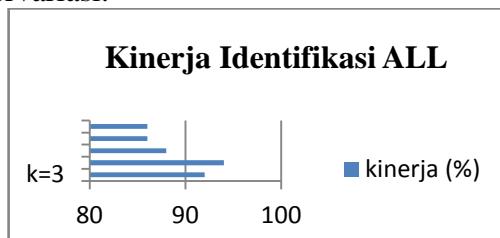
$$\text{ciri} = \begin{bmatrix} \text{rerata} \\ \text{varian} \\ \text{standardeviasi} \end{bmatrix} \dots\dots\dots (2)$$

Dari 200 data citra yang digunakan dalam, dibentuk dua matriks ciri untuk proses pelatihan dan pengujian *classifier* kNN. Matriks ciri ini berukuran 3x150 (untuk data pelatihan) dan 3x50 (untuk data pengujian). Data pelatihan diambil dari hasil ekstraksi ciri 75 sel darah putih normal dan 75 sel darah putih abnormal (ALL); sedangkan data pengujian diambil dari hasil ekstraksi ciri 25 sel darah putih normal dan 25 sel darah putih abnormal (ALL).

Klasifikasi bertujuan untuk membedakan citra leukosit normal dengan citra leukosit abnormal (ALL). Citra leukosit normal dikelompokkan dalam kelas 1 dan citra leukosit abnormal (ALL) dikelompokkan dalam kelas 2. Pada fase pelatihan, algoritma kNN melakukan penyimpanan vektor-vektor ciri dan klasifikasi dari data pelatihan yaitu berupa matriks berdimensi 3x150 yang diperoleh dari 150 citra sel darah putih berikut dengan klasifikasinya. Pada fase klasifikasi, dihitung jarak dari

vektor ciri obyek (50 data uji) yang baru terhadap seluruh vektor data pelatihan, dan diambil sebanyak k jarak yang paling dekat. Obyek yang baru pada data uji ini diklasifikasikan berdasar jumlah kelas yang terbanyak (dari k jarak terdekat yang diambil).

Dalam penelitian ini digunakan nilai k yang bervariasi yaitu 3, 5, 7, 9, dan 11. Gambar 5 memperlihatkan grafik hasil pengujian identifikasi leukemia akut menggunakan *classifier* kNN dengan nilai k yang bervariasi.



Gambar 5. Grafik kinerja kNN untuk identifikasi ALL untuk berbagai nilai k

Berdasarkan Gambar 5 maka kinerja *classifier* kNN menggunakan jarak Euclidean mencapai maksimum saat digunakan nilai $k = 5$. Nilai maksimum kinerja yang dicapai tersebut setinggi 94%; ada 3 data uji yang diklasifikasikan salah oleh *classifier* kNN. Hasil penelitian ini merekomendasikan digunakannya nilai $k = 5$ untuk identifikasi leukemia akut menggunakan *classifier* kNN. Hal yang juga menarik adalah bahwa saat digunakan nilai $k = 9$ dan $k = 11$ diperoleh capaian kinerja yang sama yaitu setinggi 86%.

4. Simpulan dan Saran

4.1. Simpulan

Menggunakan *classifier* kNN, dapat dilakukan identifikasi abnormalitas pada sel darah putih yaitu ALL atau sering disebut leukemia akut. *Classifier* kNN membedakan sel darah putih normal dan abnormal (ALL) berdasarkan tiga ciri yaitu ciri rerata, varians, dan standar deviasi yang diekstrak dari hasil segmentasi sel darah putih. Hasil penelitian menunjukkan bahwa kinerja *classifier* kNN mencapai maksimum hingga 94% dengan konstanta $k = 5$.

Menggunakan *classifier* kNN, dapat dilakukan identifikasi abnormalitas pada sel darah putih yaitu ALL atau sering disebut leukemia akut. *Classifier* kNN membedakan sel darah putih normal dan abnormal (ALL) berdasarkan tiga ciri yaitu ciri rerata, varians, dan standar deviasi yang diekstrak dari hasil segmentasi sel darah putih. Hasil penelitian menunjukkan bahwa kinerja *classifier* kNN mencapai maksimum hingga 94% dengan konstanta $k = 5$.

4.2. Saran

Pada penelitian ini dicapai kinerja *classifier* kNN setinggi 94%. Kinerja *classifier* kNN kemungkinan masih dapat ditingkatkan dengan menambah jumlah ciri yang digunakan ataupun dapat dengan cara meningkatkan kualitas segmentasi misalnya menggunakan segmentasi adaptif (dalam penelitian ini belum digunakan segmentasi adaptif).

DAFTAR PUSTAKA

- Duda, R.O., Harl, P.E., Stork, D.G., 1997, *Pattern Classification*, Second Edition, John Wiley & Sons, Inc.
- Theodoris, Sergios 2010, *An Introduction to Pattern Recognition: A Matlab Approach*, Elseivier Inc.
- Hiremath dkk, 2010, *Automated Identification and Classification of White Blood Cells (Leucocytes) in Digital Microscopic Images*, IJCA Special Issue on “Recent Trends in Image Processing and Pattern Recognition, page 59 – 63.
- Jain, Anil K., 1989, *Fundamental of Digital Image Processing*, Prentice Hall International, Inc. Singapore.
- Khashman, A, 2008, *Blood Cell Identification Using A Simple Neural Network*, International Journal of Neural Systems, Vol. 18 No. 5 (2008) page 453-458.
- Piuri, V & Scotti, F, 2004, *Morphological Classification of Blood Leucocytes by Microscope Images*, Cimsa 2004 – IEEE International Conference on Computational Intelligence for

- Measurement Systems and Applications, 14 – 16 July 2004, page 103 – 108.
- Putzu, L & Ruberto, C, 2013, *White Blood Cells Identification and Counting from Microscopic Blood Image*, World Academy of Science, Engineering and Technology, International Journal of Medical, Pharmaceutical Science and Engineering, Vol. 7 No. 1 2013, page 15 – 22.
- Scotti, F, 2005, *Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images*, CIMSA 2005 – IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, 20 – 22 July 2005, page 96 – 101.
- Sergios Theodori, 2010, An Introduction to Pattern Recognition: A Matlab Approach, Elsevier Inc.
- H. Mirzaalian dkk, 2007, *Pre-processing Algorithms on Digital Mammograms*, MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo,Japan.