

Analisis Perbandingan Metode *Support Vector Machine* (SVM) dan *Decision Tree* Pada Klasifikasi Penyakit *Acute Lymphoblastic Leukimia* (ALL)

Fahmi Alamsyah¹, Muhammad Dzikrullah Suratin.², Mustamin Hamid³

¹Teknik Informatika Universitas Muhammdiyah Maluku Utara

²Teknik Informatika Universitas Muhammdiyah Maluku Utara

³Teknik Informatika Universitas Muhammdiyah Maluku Utara

email: fahmialamsyah1010@gmail.com*; ² Irul.dzi@gmail.com; ³ hamidmustamin@gmail.com

Abstraksi

Penelitian ini mengajukan sebuah sistem untuk melakukan perbandingan metode klasifikasi *Support Vector Machine* (SVM) dan *Decision Tree* untuk mengidentifikasi penyakit *Acute Lymphoblastic Leukimia*. Algoritma penelitian yang digunakan terdiri dari beberapa tahapan : *Preprocessing* menggunakan metode median filter. *Segmentasi* menggunakan metode *K-Means Clustering* segmentasi warna digunakan agar dapat memisahkan citra sel darah putih dan objek yang lain. Hasil penelitian sebelumnya menggunakan metode *Support Vector Machine* (SVM) memperoleh tingkat akurasi sebesar 95,45%. Setelah dilakukan penelitian dengan membandingkan metode klasifikasi menggunakan *Decision Tree*, diperoleh tingkat akurasi sebesar 65,90%.

Kata Kunci : *Leukemia, K-means Clustering, Support Vector Machine, Decision Tree.*

Abstract

This study proposes a system to compare the *Support Vector Machine* (SVM) and *Decision Tree* classification methods to identify *Acute Lymphoblastic Leukemia*. The research algorithm used consists of several stages: *Preprocessing* using the median filter method. *Segmentation* uses the *K-Means Clustering* method. Color segmentation is used to separate images of white blood cells and other objects. The results of previous research using the *Support Vector Machine* (SVM) method obtained an accuracy rate of 95.45%. After conducting research by comparing the classification methods using the *Decision Tree*, an accuracy rate of 65.90% was obtained

Keywords : *Leukemia, K-means Clustering, Support Vector Machine, Decision Tree.*

PENDAHULUAN

Acute Lymphoblastic Leukimia (ALL) merupakan salah satu jenis leukimia (kanker sel darah putih) yang pada umumnya terjadi pada anak-anak. Angka kejadian tertinggi sekitar 80% penderita berumur 1-4 tahun dan keganasannya berkembang dengan cepat, jika tidak segera dirawat bisa berakibat fatal terhadap kondisi penderita dalam beberapa bulan.[1] Deteksi dini sangat penting untuk Langkah pengobatan yang lebih efektif.ALL didiagnosis dengan metode pengujian mikroskop sediaan apus darah (A peripheral blood smear)[2]. Ciri-ciri morfologi sel darah putih diamati untuk mengidentifikasi dan mengklasifikasi adanya sel-sel leukimia. Investigasi mikroskopik dilakukan secara manual oleh hematologist melalui identifikasi visual di bawah mikroskop, oleh karena itu metode ini memerlukan tenaga dan waktu yang lama serta tingkat kesalahan antara 30% dan 40% tergantung pada subyektifitas pengalaman hematologist, ketelitian dan konsentrasi [3]. Dengan adanya perkembangan di bidang pengolahan citra medis dan sistem diagnosis yang terbantu komputer, diharapkan dapat mengurangi efek subyektifitas dan membantu dokter dalam proses diagnosa. Penggunaan teknologi berbasis data di bidang medis dengan pengenalan pola mendapat perhatian lebih. Untuk mendukung hal ini, dapat digunakan Teknik data mining untuk mengenali informasi berharga dari kumpulan informasi data ALL. Klasifikasi adalah salah satu jenis dari supervised learning yang dimana supervised learning merupakan tipe pembelajaran dari data mining.

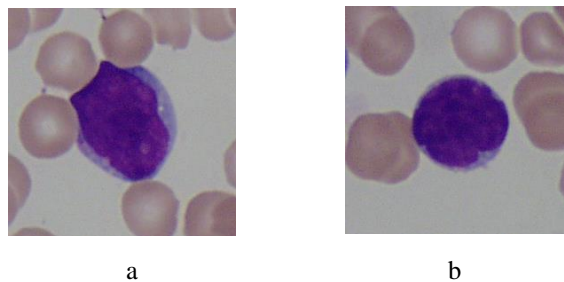
Pada beberapa penelitian sebelumnya telah dilakukan implementasi untuk memprediksi penyakit leukimia tipe ALL. Salah satu penelitian yang dilakukan pada tahun 2015 oleh M Dzikrullah S., dkk melakukan Identifikasi Sel *Acute Lymphoblastic Leukemia* (ALL) pada citra *Peripheral Blood Smear* Berdasarkan Morfologi Sel Darah Putih, metode SVM (*Support Vector Machine*) digunakan dalam tahap klasifikasinya, menghasilkan tingkat akurasi sebesar 95,45%.[4] Penelitian lain, dilakukan pada tahun 2021 oleh Shofiya dan Bagus melakukan Identifikasi *Acute Lymphoblastic Leukemia* Pada Citra Mikroskopis Menggunakan Algoritma *Naïve Bayes* menghasilkan tingkat akurasi sebesar 80%.[5] Banyak algoritma data mining yang telah diterapkan dan dilakukan

perbandingan dalam pengenalan pola di bidang medis seperti yang dilakukan pada tahun 2020 oleh Shidqi Aqil Naufal dan kawan-kawan yang melakukan Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk Deteksi Kanker Dengan Data Microarray, menghasilkan tingkat akurasi sebesar 98,54% pada data leukimia dengan PLS-KNN, 100% pada data lung dengan KNN, 66,52% pada data breast dengan PLS-KNN, dan 66,52% pada data colon dengan PLS-SVM. KNN mampu mendapatkan akurasi terbaik di tiga data dari empat data yang diuji. [6] Penelitian lain dilakukan pada tahun 2020 oleh Agus Heri Yunial yang melakukan Analisa Perbandingan Algoritma Klasifikasi Support vector Machine, Decision Tree dan Naïve Bayes, menghasilkan tingkat akurasi sebesar 87,43 pada algoritma Decision Tree, 87,16% pada algoritma Support vector Machine, dan 84,92% pada algoritma Naïve Bayes. [7] Penelitian lain juga dilakukan oleh Joshua Bonardo junior dan kawan-kawan yang melakukan Perbandingan Akurasi Algoritma Decision Tree dan Support vector Machine Pada Penyakit Diabetes, menghasilkan tingkat akurasi sebesar 85,28% pada metode Decision Tree dan 83,85% pada algoritma Support vector Machine. [8]

Dalam penelitian ini dilakukan perbandingan terhadap dua metode klasifikasi antar metode *Support Vector Machine* (SVM) dan *Decision Tree* untuk mengetahui tingkat akurasi pada klasifikasi penyakit *Acute Lymphoblastic Leukimia* (ALL).

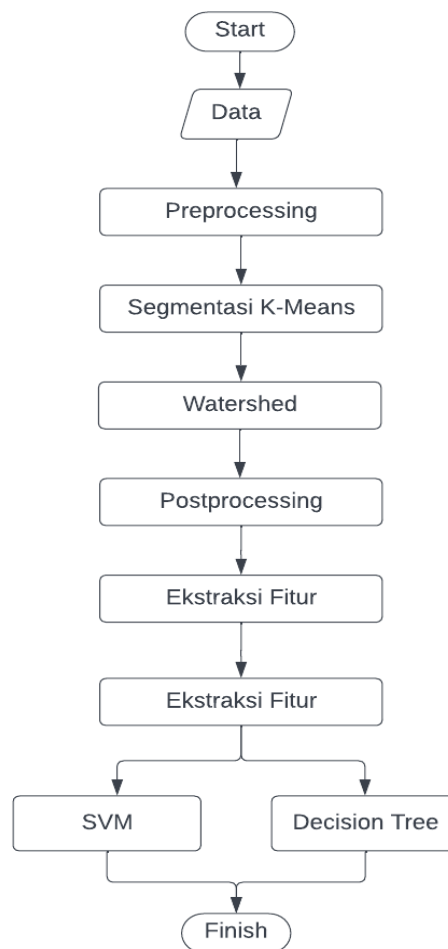
METODE PENELITIAN

Data berupa citra sel darah putih didapatkan dari M.Dzikrullah S. yang terdiri dari 214 data citra sel darah putih yang terdiri dari dua versi yaitu versi ALL_IDB1 yang merupakan kumpulan data citra yang diambil menggunakan kamera Canon PowerShot G5 dengan perbesaran mikroskop antara 300-500 kali yang berukuran 2592x1944 pixel 24 bit dengan format JPG dan versi ALL_IDB2 yang merupakan hasil *cropped* ROG (*Region of Interest*) citra sel darah putih normal dan *Acute Lymphoblastic Leukimia* yang berukuran 257x257 pixel dari ALL_IDB1. Gambar 1 menunjukkan sampel data yang digunakan.



Gambar 1. Sampel Citra
(a) Sel darah ALL
(b) Sel darah Normal

Dalam penelitian ini terdapat beberapa Langkah yang harus dilakukan diantaranya : Input citra, preprocessing, segmentasi, watershed, postprocessing, ekstraksi fitur, dan klasifikasi. Proses klasifikasi menggunakan dua perbandingan metode yaitu *Support Vector Machine* (SVM) dan *Decision Tree*. dapat dilihat pada gambar 2.



Gambar 2. Alur penelitian

A. Preprocessing

Preprocessing merupakan tahapan yang penting dalam penelitian ini. Prerocessing bertujuan untuk memperbaiki kualitas citra. Pada penelitian ini tahapan preprocessing menggunakan metode median filter. Median filter adalah salah satu metode yang dapat digunakan untuk menghilangkan noise pada citra dalam domain special [9].

B. Segmentasi

Segmentasi citra bertujuan untuk memisahkan antar objek yang termuat dalam satu citra. Pada penelitian ini tapan segmentasi menggunakan *K-Means clustering* berdasarkan warna. Berdasarkan data yang diperoleh, citra sel darah putih memiliki empat warna yaitu warna putih pada latar belakang, inti sel WBC (*White Blood Cells*), yang berwarna ungu, warna biru pada sitoplasma WBC dan warna merah pada RBC (*Red Blood Cells*). Metode *K-Means clustering* mampu memisahkan komponen-komponen di atas berdasarkan warnanya.

C. Watershed

Watershed merupakan pendekatan untuk segmentasi. Konsep dari transformasi watershed adalah menganggap bahwa sebuah citra adalah 3 dimensi. Prinsip dari transformasi watershed adalah mencari garis watershed. Proses pada tahapan watershed ini yaitu citra hasil *clustering* ditrasformasi menjadi citra *gray level*, kemudian dirubah dalam format binary dengan *global thresholding* menggunakan metode otsu. Setelah itu barulah diaplikasikan operasi morfologi seperti *region filling* untuk menutup daerah objek yang berlubang, operasi erosi untuk menghilangkan piksel-piksel yang bukan termasuk inti sel kemudian disempurnakan dengan operasi dilasi.

D. Postprocessing

Tahap postprocessing ini digunakan untuk memperhalus, menajamkan, dan memperbaiki kualitas dari hasil proses segmentasi. Metode yang digunakan dalam tahap ini yaitu operasi morfologi Opening dan Closing. Operasi Opening digunakan untuk membuat kontur gambar lebih jelas, halus, dan mengurangi noise hasil segmentasi. Operasi Opening secara sederhana dilakukan dengan Operasi Erosi diikuti oleh Dilasi

E. Ekstraksi Fitur

Ekstraksi fitur merupakan tahap paling penting dalam penelitian ini. Ekstraksi ciri adalah mengestimasi ciri objek berdasarkan parameter antara satu dengan yang lain. Pada penelitian ini, ada beberapa parameter yang harus d ekstrak berdasarkan ciri morfologi sel darah putih diantaranya: area, perimeter, equiv diameter, form factor/roundness dan compactness.

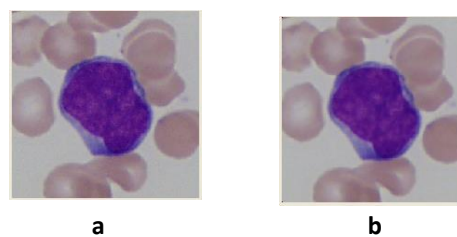
F. Klasifikasi

Pada tahap ini dilakukan perbandingan pengenalan yang terdiri dari pembelajaran dan pengujian terhadap sel leukosit normal dan ALL menggunakan *Support Vector Machine (SVM)* dan *Decision Tree*. Metode *Support Vector Machine (SVM)* bertujuan menemukan fungsi pemisah (*hyperline*) dengan solusi global optimal [10]. *Hyperline* terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin* antara kedua *hyperline* pendukung dari kedua kelas tersebut dengan mencari titik maksimalnya. Subset dari data pembelajaran yang paling informatif adalah hasil akhir dari proses pembelajaran yang disebut sebagai *support vector*. Sedangkan *Decision Tree* diterapkan untuk mempelajari klasifikasi dan memprediksi pola dari data dan menggambarkan relasi dari variabel atribut x dan variabel target y dalam bentuk pohon. *Decision tree* merupakan struktur yang menyerupai flowchart dimana untuk setiap internal node merupakan pengujian terhadap variabel atribut, tiap cabangnya adalah hasil dari pengujian tersebut, sedangkan node terluar yaitu leaf menjadi labelnya.

HASIL DAN PEMBAHASAN

A. Preprocessing

Pada tahap *preprocessing*, perbaikan kualitas citra input RGB (*Red, Green, Blue*) dilakukan per-kanal warna untuk menghilangkan noise pada citrayang disebabkan oleh proses selama pengambilan citra dengan menggunakan metode *median filter*. *Median filter* yang digunakan berukuran matrix 3×3 . Untuk pengukuran hasil kualitas citra yang dihasilkan oleh metode *median filter* yang lebih obyektif, dapat dihitung dengan menggunakan pengukuran statistic seperti MSE (*Mean Square Error*) dan PSNR (*Peak Signal to Noise Ratio*). Perbandingan hasil sebelum dan setelah dilakukan *median filter* dapat dilihat pada gambar 3.



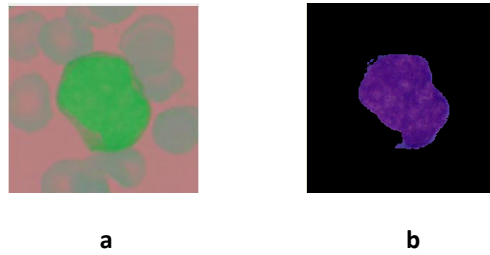
Gambar 3. Hasil median filter

(a) Citra asli

(b) Citrah setelah median filter

B. Segmentasi

Pada tahapan segmentasi terdapat beberapa Langkah yang harus dilakukan untuk mendapat kan objek citra inti sel. Tahap pertama yaitu citra hasil *median filter* dikonversi dari ruang warna RGB ke ruang warna L^*a^*b atau dikenal dengan nama CIELab. Ruang warna CIELab dipilih karena menyediakan spectrum cahaya yang lengkap untuk mata manusia, ditetapkan oleh Komisi Nasional tentang iluminasi warna. Komponen L memuat unsur kecerahan dari suatu warna, komponen a menampilkan warna sepanjang sumbu merah dan hijau dan komponen b menampilkan warna sepanjang sumbu warna biru dan kuning. Tahapan selanjutnya yaitu melakukan segmentasi citra dengan menggunakan metode *K-means clustering*. Metode *K-Means clustering* digunakan karena mampu memisahkan objek berdasarkan warna. Hasil konversi warna RGB ke ruang warna L^*a^*b dan segmentasi *K-Means clustering* dapat dilihat pada Gambar 4.



Gambar 4. Hasil segmentasi
(a) Citra CIELab
(b) Citra segmentasi *K-means clustering*

C. Watershed

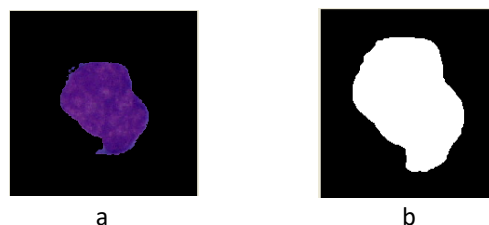
Pada tahapan ini citra hasil segmentasi menggunakan *K-Means clustering* di transformasi menjadi citra *gray level*, kemudian dirubah dalam format binary dengan menggunakan *global thresholding* dengan menggunakan metode otsu. Setelah dilakukan transformasi *gray level*, dan dirubah dalam format binary, barulah diaplikasikan operasi morfologi seperti *region filling*, untuk menutupi daerah pada citra yang berlubang, operasi erosi yang berfungsi untuk memisahkan objek yang bukan merupakan inti sel. Hasil citra segmentasi dan *watershed* dapat dilihat pada gambar 5.



Gambar 4. Hasil watershed
(a) Citra segmentasi *K-means clustering*
(b) Citra *watershed*

D. Postprocessing

Tahapan *postprocessing* pada penelitian ini digunakan untuk memperhalus mempertajam dan memperbaiki kualitas citra yang dihasilkan dari tahapan segmentasi. Metode yang digunakan dalam tahap ini yaitu operasi morfologi Opening dan Closing. Operasi Opening digunakan untuk membuat kontur gambar lebih jelas, halus, dan mengurangi noise hasil segmentasi. Operasi Opening secara sederhana dilakukan dengan Operasi Erosi diikuti oleh Dilasi sehingga citra yang dihasilkan terlihat lebih sempurna setelah melewati tahap ini. Hasil dari tahapan *postprocessing* dapat dilihat pada gambar 6.



Gambar 4 Hasil *postprocessing*
(a) Citra segmentasi *K-means clustering*
(b) Citra *postprocessing*

E. Ekstraksi Fitur

Berdasarkan penjelasan yang telah dipaparkan sebelumnya, maka hasil citra yang telah melalui beberapa tahapan sebelumnya kemudian diekstrak cirinya. Pada bagian ini nucleus diektrak dengan karakteristik morfologi. Segala informasi yang berhubungan dengan ukuran betuk dari inti sel darah putih termasuk dalam karakteristik morfologi. Karakteristik yang berhubungan dengan ukuran pada penelitian ini meliputi *area*, *parameter*, dan *diameter*. Sedangkan *roundness* dan *compactness* termasuk dalam ciri bentuk.

F. Klasifikasi

Berdasarkan data hasil segmentasi yang diperoleh, maka jumlah citra sel darah putih yang digunakan untuk proses klasifikasi yaitu sebanyak 214 citra yang terdiri dari 123 citra sel darah ALL dan 91 citra sel darah normal. Dari jumlah data keseluruhan yang diperoleh, data citra dibagi menjadi data pembelajaran dan data pengujian dengan jumlah data pembelajaran sebanyak 170 citra dan data pengujian sebanyak 44 citra.

Dalam penelitian ini, proses klasifikasinya menggunakan dua metode klasifikasi untuk dilakukan perbandingan yaitu metode *Support Vector Machine* (SVM) dan *Decision Tree*. Setelah dilakukan pelatihan terhadap 170 citra dan pengujian terhadap 44 citra menggunakan kedua metode, maka hasil yang diperoleh dapat dilihat pada tabel 1 :

Tabel 1. Hasil Pengujian *Support Vector Machine* (SVM) dan *Decision Tree*

Metode	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
<i>Support Vector</i>	24	18	1	1
<i>Decision tree</i>	13	16	12	2

Berdasarkan tabel hasil pengujian data diatas, maka dapat dihitung sensitifitas, spesififikasi dan tingkat akurasi menggunakan persamaan sebagai berikut :

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 = \frac{24}{24 + 1} \times 100 = 96\%$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100 = \frac{18}{18 + 1} \times 100 = 94,73\%$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 = 95,45\%$$

Perhitungan *Support Vector Machine*

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 = \frac{13}{13 + 2} \times 100 = 86,66\%$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100 = \frac{16}{16 + 12} \times 100 = 57,14\%$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 = 65,90\%$$

Perhitungan *Decision Tree*

Berdasarkan perhitungan diatas maka diperoleh hasil yang telah di buat dalam bentuk tabel sebagai berikut :

Tabel 2. Hasil Perhitungan *Support Vector Machine (SVM)* dan *Decision Tree*

Metode	<i>Sensitifity</i>	<i>Specificity</i>	<i>Accuracy</i>
<i>Support Vector</i>	96%	94,73%	95,45%
<i>Decision tree</i>	86,66%	57,14%	65,90%

KESIMPULAN

Hasil perbandingan antara metode klasifikasi *Support Vector Machine (SVM)* dan *Decision Tree* dalam mengidentifikasi penyakit *Acute Lymphoblastic Leukimia (ALL)* dengan menggunakan 170 sampel darah untuk data training dan 44 sampel darah untuk data testing, peneliti sebelumnya telah melakukan penelitian sejenis dengan menggunakan metode *Support Vector Machine (SVM)* dan memperoleh hasil akurasi sebesar 95,45%. Pada penelitian ini peneliti melakukan perbandingan dengan menggunakan metode *Decision Tree* menghasilkan tingkat akurasi sebesar 65,90%. Hasil akurasi yang didapatkan dengan menggunakan metode *Decision Tree* jauh lebih kecil jika dibandingkan dengan metode *Support Vector Machine (SVM)*.

DAFTAR PUSTAKA

- [1] American Cancer Society, 2014, Leukemia: Acute Lymphocytic Overview. www.cancer.org, diakses senin 28 maret 2015.
- [2] Putzu L., Di Ruberto C., 2013, White Blood Cells Identification and Counting from Microscopic Blood Image. International Journal of Medical, Health, Biomedical and Pharmaceutical Engineering Vol:7, No:1.
- [3] Reta C., dkk., 2010, Segmentation of Bone Marrow Cell Images for Morphological Classification of Acute Leukemia. Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference, USA.
- [4] Suratin M. D., 2015, Identifikasi Sel Acute Lymphoblastic Leukemia (ALL) pada Citra Peripheral Blood Smear Berdasarkan Morfologi Sel Darah Putih. elektronik Jurnal Arus Elektro Indonesia (eJAEI), 7–12.
- [5] Syidada S., 2021, Identifikasi Acute Lymphoblastic Leukemia pada Citra Mikroskopis Menggunakan Algoritma Naïve Bayes. Journal of Science and Technology, 78-83
- [6] Naufal, Shidqi, Aqil., 2020, Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk Deteksi Kanker dengan Data Microarray. JURIKOM (Jurnal Riset Komputer), 162-168.
- [7] Yunial, Heri, Agus., 2020, Analisa Perbandingan Algoritma Klasifikasi Support Vector Machine, Decision Tree dan Naïve Bayes. Prosiding Seminar Nasional Informatika dan Sistem Informasi, 138-156
- [8] Junior, Joshua, Bonardo., 2021, Perbandingan Akurasi Algoritma Decision Tree dan Algoritma Support Vector Machine Pada Penyakit Diabetes. e-Proceeding of Engineering. 9749-9756.
- [9] Prasetyo Eko, 2011, Pengolahan Citra Digital dan Aplikasinya Menggunakan Matlab. Yogyakarta: ANDI.
- [10] Nugroho A. S., Witarto A. B., Handoko D., 2003, Application of Support Vector Machine in Bioinformatics. Proceeding of Indonesian Scientific Meeting in Central Japan.

Biodata Penulis

Fahmi Alamsyah, memperoleh gelar Sarjana Komputer (S.Kom), Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Maluku Utara, lulus tahun 2022. Saat ini belum bekerja