

# Perbandingan Distance Metric pada Nearest Neighbour untuk Klasifikasi Sel Darah Putih

<sup>1</sup>Felix Indra Kurniadi, <sup>2</sup>Vinnia Kemala Putri

<sup>1</sup>School of Engineering and Technology: Informatics Engineering, Tanri Abeng University, Jakarta, Indonesia

<sup>2</sup>Fakultas Ilmu Komputer, Universitas Indonesia, Depok, Indonesia

felixindra@tau.ac.id

vinnia.kemala51@ui.ac.id

Diterima 18 Oktober 2019

Disetujui 24 Juni 2019

**Abstract**—White blood cells, have a function to protect human body from viruses, bacteria and any other harmful substance. In this research, we increased our dataset using augmentation techniques such as rotation. While the other image pre-processing techniques that we used is resize window-size of the images into 40x40 pixel and 50x50 pixel. Local Binary Pattern was proposed for feature extraction using Euclidean distance, Chebyshev distance and Minkowski distance as classifier. The result showed that using Euclidean and Minkowski distance give same result and also the best result compared to Chebyshev distance

**Index Terms**—Chebyshev distance, Euclidean distance, Minkowski distance, Local Binary Pattern, white blood cells

## I. PENDAHULUAN

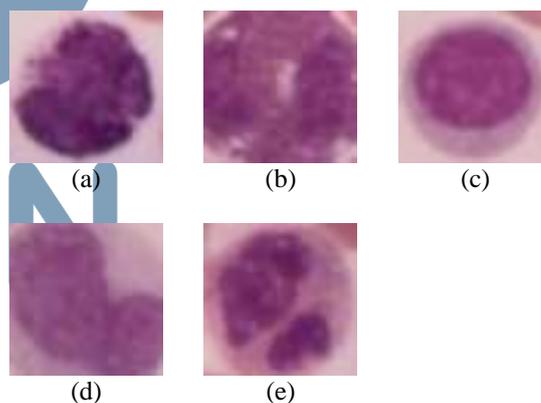
Darah merupakan salah satu bagian terpenting dalam tubuh manusia, tidak seorangpun yang bisa hidup tanpa memiliki darah. Darah memiliki tiga peranan penting yaitu sebagai transportasi, regulasi dan proteksi. Darah berperan dalam melakukan transportasi nutrient dan oksigen ke seluruh tubuh, membuang produk sisa, meregulasikan pH tubuh dan berperan sebagai pelindung dari serangan virus, bakteri maupun mikro organisme lainnya [1].

Di dalam tubuh manusia sendiri terdapat tiga komponen darah yaitu eritrosit atau sel darah merah, leukosit atau sel darah putih dan trombosit atau keping darah. Darah putih dipilih dalam penelitian ini dikarenakan memiliki peranan untuk melindungi tubuh manusia dari berbagai serangan bakteri, virus dan mikro organisme yang berbahaya bagi tubuh manusia. Sayangnya bentuk sel darah putih beragam sehingga dibutuhkan proses inspeksi dan pengenalan yang lebih baik, sehingga memudahkan para hematologis untuk mendiagnosis beberapa penyakit dalam darah seperti AIDS, leukemia dan kanker darah [2].

Pengenalan terhadap darah putih saat ini membutuhkan waktu yang lama dikarenakan proses pengenalan dilakukan secara manual. Proses ini membutuhkan waktu yang lama, para ahli yang bekerja dibidang ini dan subjektifitas yang terjadi dikarenakan *human error*. Sayangnya proses pengenalan sel darah

otomatis hanya dapat dilakukan oleh rumah sakit besar, sedangkan bagi rumah sakit kecil atau puskesmas di daerah pedesaan hanya dapat dilakukan secara manual [3].

Sel darah putih memiliki karakteristik yang berbeda untuk setiap tipenya. Keunikan setiap sel darah putih inilah yang membuat sel darah putih butuh untuk diklasifikasikan. Proses pengklasifikasian akan dilakukan dengan menggunakan citra darah yang diambil dari mikroskop dan diambil citranya menggunakan kamera *webcam*.



Gambar 1. Lima Tipe Sel Darah Putih: a) basofil, b) eosinofil c) limfosit d) monosit e) neutrofil

Penelitian ini akan menggunakan metode *Local Binary Pattern* sederhana dengan menggunakan *K-Nearest Neighbor* dengan tiga *distance metric* yang berbeda. Tiga *distance metric* yang digunakan adalah *Chebyshev distance*, *Euclidean distance* dan *Minkowski distance*.

Artikel ilmiah ini dibagi menjadi 5 bagian dimana Bagian 1 adalah pendahuluan, Bagian 2 merupakan penelitian sebelumnya, Bagian 3 eksperimen, Bagian 4 hasil eksperimen dan Bagian 5 kesimpulan dan penelitian selanjutnya.

## II. PENELITIAN SEBELUMNYA

Beberapa penelitian yang coba mengatasi problem didalam melakukan pengklasifikasian sel darah putih. Lina mencoba mengatasi permasalahan ini dengan mengajukan metode Scale Invariant Feature Transform dengan menggunakan classifier Co-occurrence matrix untuk citra warna dan Euclidean distance untuk fitur tekstur. Penelitian yang dilakukan ini memberikan kesimpulan bahwa pencahayaan dan *window size* memberikan perbedaan hasil yang signifikan [5].

Sholeh mengajukan segmentasi terhadap sel darah putih menggunakan Median Filtering dan Otsu threshold, hasil dari penelitian yang dilakukan memberikan kesimpulan bahwa metode yang digunakan dapat melakukan segmentasi terhadap nucleus dan sitoplasma [3],

Tabrizi mengajukan metode untuk melakukan pengenalan sel darah putih dengan menggunakan PCA dan LVQ. Metode ini memperbaiki *running time* dari metode sebelumnya akan tetapi akurasi yang dihasilkan tidak jauh berbeda dengan metode sebelumnya [2].

Gautam mengajukan pengenalan sel darah menggunakan metode Morphological Feature. Metode yang diusulkan ini memberikan hasil terbaik pada pengenalan limfosit dengan akurasi yang mencapai 100% untuk limfosit [6].

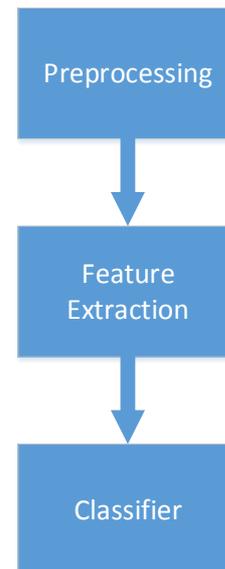
Kurniadi mengajukan pengklasifikasian sel darah putih dengan menggunakan berbagai metode LBP yaitu *Local Binary Pattern*, *Local Binary Pattern Uniform*, *Local Binary Pattern Rotation Invariant* dan *Local Binary Pattern Rotation Invariant Uniform*. Hasil yang didapatkan menyebutkan bahwa metode LBP-RI dan LBP-RIU memberikan hasil yang baik pada kedua *windows size* dengan menggunakan *K-NN classifier* dengan *distance metric* menggunakan *Euclidean distance* [7].

Berdasarkan penelitian yang dilakukan oleh [7], yang menyatakan bahwa penggunaan menggunakan *K-NN classifier* pada *distance metric Euclidean distance* memberikan hasil yang paling baik dibandingkan dengan *classifier* lainnya. Penulis ingin membuktikan penggunaan *distance metric* lainnya yang dapat memperbaiki akurasi atau menghasilkan hasil yang setara dengan penggunaan *Euclidean distance*.

## III. METODOLOGI PENELITIAN

Metodologi penelitian ini dapat digambarkan dengan bagan seperti pada **Gambar 2**. Tiga tahapan yang dilakukan adalah tahap *pre-processing*, kemudian dilanjutkan dengan tahapan pencarian fitur dari citra digital kemudian akan dilakukan pembuatan model dengan menggunakan metode *K-Nearest Neighbor* dengan tiga *distance metric* yang berbeda.

Data yang digunakan adalah data citra yang merupakan dataset sel darah putih dari Universitas Tarumanagara.



**Gambar 2.** Skema Eksperimen

### A. Tahapan Preprocessing

Pada tahapan ini akan dilakukan dua tahapan yaitu pemotongan manual terhadap citra digital, perubahan citra digital menjadi 40x40 window size dan 50 x 50 window size dan proses augmentasi citra.

Pada tahapan pemotongan citra, citra darah dipotong manual dan dipisahkan berdasarkan kelas yang sesuai dalam hal ini adalah dibagi menjadi: eritrosit, limfosit, eosinophil, basophil, dan neutrophil. Hasil yang didapatkan dari proses ini adalah terdapat sebanyak 24 citra basophil, 93 citra eosinophil, 69 citra limfosit, 24 citra monosit dan 94 citra neutrophil.

Tahapan berikutnya adalah melakukan perubahan *windows size* menjadi 40 x 40 dan 50 x 50 dari citra darah yang sudah didapatkan pada tahapan sebelumnya. Proses ini dilakukan berdasarkan [5], bahwa penggunaan *windows size* berpengaruh pada hasil pengklasifikasian. Proses perubahan *window size* dilakukan untuk mendapatkan nilai *window size* yang sejenis sehingga dapat mengabaikan resolusi dalam penelitian ini.

Tahapan terakhir adalah tahapan augmentasi citra, tahapan augmentasi citra dilakukan adalah untuk memperbanyak citra, proses ini dilakukan agar tidak adanya unbalance data dan memperbanyak data untuk menghasilkan model yang lebih baik dibandingkan model yang menggunakan data citra sesuai dengan yang didapatkan. Proses augmentasi dilakukan dengan cara melakukan rotasi terhadap citra utamanya. Hasil yang didapatkan sebanyak 500 citra untuk setiap citra jenis darah putih.

### B. Ekstraksi Ciri

Tahapan ini akan melakukan pengekstraksian ciri menggunakan LBP sederhana, dimana tidak menggunakan konsep *rotation invariant*, *rotation invariant uniform* dan *uniform*. Hal ini dilakukan untuk

mempbandingkan hasil terbaik dari *distance metric* terhadap sebuah ekstraksi fitur LBP sederahana.

Tahapan-tahapan yang dilakukan dalam menggunakan metode LBP adalah:

1. Mencari LBP operator dengan fungsi, penelitian ini akan menggunakan ketetanggan 8 yaitu dari satu piksel hanya melihat ketetanggaannya. Persamaan 1 dan Persamaan 2 merupakan persamaan LBP[8]:

$$LBP_{p,r} = \sum_0^{p-1} s(g_p - g_c)2^p \quad (1)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

$P$  adalah jumlah tetangga pixel,  $R$  adalah radius,  $g_p$  adalah nilai keabuan tetangga,  $g_c$  adalah nilai keabuan dari pusat piksel,  $s(x)$  adalah fungsi mencari nilai biner berdasarkan pixel tetangga dan  $2^p$  adalah fungsi merubah nilai biner kedesimal.

2. Setelah mendapat nilai dari LBP operator kemudian dibuat histogramnya dari nilai tersebut sehingga dari sebuah citra akan didapatkan 0-255 fitur.

### C. Klasifikasi

Proses pengklasifikasian sel darah putih akan menggunakan metode *K-Nearest Neighbor* dengan menggunakan tiga *distance metric* yang berbeda yaitu *Euclidean distance*, *Minkowski distance* dan *Chebyshev distance*.

Metode *Euclidean distance* memiliki persamaan yang dapat dilihat pada Persamaan 3[9]:

$$d(x, y) = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} \quad (3)$$

Metode *Minkowski distance* memiliki persamaan yang dapat dilihat pada persamaan 4 [10]:

$$d(x, y) = \left\{ \sum_{k=1}^n |x_k - y_k| \right\}^{\frac{1}{r}} \quad (4)$$

Metode *Chebyshev distance* memiliki persamaan yang dapat dilihat pada persamaan 5:

$$d(x, y) = \max\{|x_j - y_j|\} \quad (5)$$

Dimana  $d(x,y)$  adalah nilai dari jarak dua nilai,  $r$  adalah banyaknya parameter dan  $x,y$  adalah titik jarak.

### D. Pengukuran Evaluasi

Pada penelitian ini ada beberapa kondisi yang akan digunakan, salah satunya proses pelatihan dan pengujian akan menggunakan konsep 80% data untuk proses pelatihan dan 20 % data akan digunakan untuk proses pengujian. Proses ini akan dilakukan secara acak.

Proses pengukuran evaluasi dari penelitian ini akan menggunakan akurasi dan Cohen Kappa:

$$acc = \frac{tp + tn}{fp + fn + tp + tn} \quad (6)$$

dimana  $tp$  dan  $tn$  merupakan jumlah klasifikasi yang benar.

Sedangkan *Cohen Kappa* statistik digunakan sebagai evaluasi dikarenakan metode ini sering digunakan untuk mengukur *interatter reliability*[11].

$$kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (7)$$

dimana  $Pr(a)$  merupakan total probabilitas kesepakan yang terobservasi dan  $Pr(e)$  total probabilitas yang diharapkan.

### IV. HASIL EKSPERIMEN

Pada penelitian ini akan citra sel darah akan melalui proses pengeksraksian ciri dengan metode LBP, setelah itu akan dilakukan proses pencarian model dengan berbagai *distance metric* yang diajukan (*Euclidean*, *Minkowski*, *Chebyshev distance*). Hasil dari Tabel akurasi dan *Cohen Kappa* untuk citra dengan ukuran 50x50 dapat dilihat pada **Tabel 2** dan **Tabel 3** dan, akurasi dan *Cohen Kappa* untuk citra dengan ukuran 40 x 40 dapat dilihat pada **Tabel 4** sampai **Tabel 5**.

**Tabel 2.** Hasil akurasi berbagai metode *distance metric* pada citra dengan *window size* 50 x 50

Metode	Akurasi (%)
<i>Euclidean</i>	64.2
<i>Chebyshev</i>	63.4
<i>Minkowski</i>	64.2

**Tabel 3.** Hasil Cohen Kappa's berbagai metode *distance metric* pada citra dengan *window size* 50 x 50

Metode	Cohen kappa (%)
<i>Euclidean</i>	55.18
<i>Chebyshev</i>	54.26
<i>Minkowski</i>	55.18

**Tabel 4.** Hasil akurasi berbagai metode *distance metric* pada citra dengan *window size* 40 x 40

Metode	Akurasi (%)
<i>Euclidean</i>	58.6
<i>Chebyshev</i>	52
<i>Minkowski</i>	58.6

**Tabel 5.** Hasil Cohen Kappa's berbagai metode *distance metric* pada citra *window size* 40 x 40

Metode	Cohen kappa (%)
<i>Euclidean</i>	48.07
<i>Chebyshev</i>	39.85
<i>Minkowski</i>	48.07

Berdasarkan hasil akurasi dan nilai *cohen kappa* yang didapatkan maka didapatkan hasil dengan penggunaan *distance metric* dengan metode *Euclidean distance* dan *Minkowski distance* memberikan hasil yang sama dan Jika dibandingkan dengan keseluruhan *distance metric* maka *Chebyshev* memberikan hasil terburuk untuk setiap kasus.

Selain dari akurasi yang didapatkan, dapat dilihat bahwa setiap *window size* memberikan pengaruh terhadap hasil akurasi dan nilai dari cohen kappa. Dimana hasil terbaik diberikan oleh *window size* dengan ukuran 50x50. Hal ini sesuai dengan penelitian yang dilakukan oleh [7].

Jika dilihat pada hasil diatas dapat dilihat bahwa nilai yang dihasilkan tidak melebihi 70%. Hal ini disebabkan ekstraksi fitur dari citra yang digunakan menggunakan ekstraksi ciri yang sederhana. Akan tetapi dengan menggunakan metode ekstraksi yang lebih baik seperti digunakan pada [7]

#### V. SIMPULAN DAN PENELITIAN SELANJUTNYA

Pada penelitian untuk membandingkan ketiga *distance metric* yang tepat digunakan dalam penggunaan metode *K-Nearest Neighbor* memberikan hasil bahwa penggunaan *Euclidean distance* dan *Minkowski distance* memberikan hasil yang sama dan lebih baik jika dibandingkan dengan menggunakan *Chebyshev's Distance*.

Perbedaan *window size* pada KNN tidak memberikan hasil yang begitu memuaskan sehingga pada penelitian selanjutnya sebaiknya *window size* 40 x 40 tidak digunakan pada klasifier KNN.

Berdasarkan temuan ini dapat disimpulkan bahwa penggunaan *window size* untuk metode LBP sangat mempengaruhi hasil dan *distance metric* yang baik digunakan untuk pengklasifikasian citra darah adalah *Euclidean* dan *Minkowski*.

#### UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terimakasih untuk Gunawan dari Universitas Tarumanagara untuk memberikan dataset dan feedback untuk penelitian ini

#### DAFTAR PUSTAKA

- [1] myVMC, "Blood Function and Composition," 2017. [Online]. Available: <https://www.myvmc.com/anatomy/blood-function-and-composition/>. [Accessed: 08-Mar-2018].
- [2] P. R. Tabrizi, S. H. Rezatofighi, and M. J. Yazdanpanah, "Using PCA and LVQ Neural Network for Automatic Recognition of Five Types of White Blood Cells," in *32nd Annual International Conference of the IEEE EMBS*, 2010, pp. 5593–5596.
- [3] F. . Sholeh, "White Blood Cells Segmentation for Fresh Blood Smear Images," in *The 5th International Conference on Advance Computer Science and Information System*, 2013.
- [4] J. Rawat, H. . Bhadauria, A. Singh, and J. Virmani, "Review of Leukocyte Classification Techniques for Microscopic Blood Images," in *2nd International Conference on Computing for Sustainable Global Development*, 2015.
- [5] Lina, A. Chris, B. Mulyawan, and A. B. Dharmawan, "A Leukocyte Detection System Using Scale Invariant Feature Transform Method," *Int. J. Comput. Theory Eng.*, vol. 8, no. 1, pp. 69–73, 2016.
- [6] A. Gautam and H. Bhadaria, "Classification of White Blood Cells Based on Morphological Features," in *2014 International Conference on Advances in Computing, Communications and Informatics*, 2014.
- [7] F. I. Kurniadi, "Perbandingan Local Binary Pattern untuk Klasifikasi Sel Darah Putih," *Ultimatics*, vol. IX, no. 2, pp. 118–121, 2017.
- [8] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
- [9] K. N. Stevens, T. M. Cover, and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [10] R. Kamimura and O. Uchida, "Greedy network-growing by Minkowski distance functions," in *2004 IEEE International Joint Conference on Neural Networks*, 2004.
- [11] S. M. Vieira, U. Kaymak, and J. M. C. Sousa, "Cohen 's Kappa Coefficient as a Performance Measure for Feature Selection," in *International Conference on Fuzzy Systems*, 2010.