

Analisis Performa Metode Klasifikasi Dataset Multi-Class Kanker Kulit Menggunakan KNN dan HOG

¹Sarwati Rahayu, ²Sulis Sandiwarno, ³Erwin Dwika Putra, ⁴Marissa Utami, ⁵Hadiguna Setiawan

^{1,2}Fakultas Ilmu Komputer, Universitas Mercu Buana, Indonesia

^{3,4,5}Research and Development, ESC Technology, Indonesia

^{3,4}Fakultas Teknik, Universitas Muhammadiyah Bengkulu, Indonesia

¹sarwati@mercubuana.ac.id, ²sulis.sandiwarno@mercubuana.ac.id, ³erwindwikap@gmail.com,

⁴marissautami@umb.ac.id, ⁵hadiguna.setiawan19@gmail.com

Article Info

Article history:

Received, 2024-05-15

Revised, 2024-06-03

Accepted, 2024-06-07

Kata Kunci:

KNN

HOG

skin cancer

ABSTRAK

Deteksi kanker kulit pada fase awal merupakan tantangan bahkan bagi dokter kulit. Penelitian ini bertujuan untuk menganalisis performa metode klasifikasi pada dataset multiclass kanker kulit menggunakan K-nearest neighbor (KNN) dan histogram of oriented gradients (HOG). Dataset diambil secara publik dengan nama dataset Skin Cancer MNIST: HAM10000 dataset berjumlah 10,015 data. Eksperimen pertama menggunakan dengan parameter pixels per cell yaitu 8,8 dan cells per block yaitu 2,2 mendapatkan akurasi 60,58%. Eksperimen kedua menggunakan dengan parameter pixels per cell yaitu 8,8 dan cells per block yaitu 2,2 mendapatkan akurasi 60,58%. Eksperimen terakhir menggunakan dengan parameter pixels per cell yaitu 8,8 dan cells per block yaitu 2,2 mendapatkan akurasi paling baik yaitu 61,43%.

ABSTRACT

Detection of skin cancer in its early phase is a challenge even for dermatologists. This study aims to analyze the performance of classification methods on multiclass skin cancer datasets using K-nearest neighbor (KNN) and histogram of oriented gradients (HOG). The dataset is taken publicly under the name Skin Cancer MNIST dataset: HAM10000 dataset totaling 10,015 data. The first experiment used the pixels per cell parameter of 8.8 and cells per block of 2.2 to get an accuracy of 60.58%. The second experiment used the pixels per cell parameter of 8.8 and cells per block of 2.2 to get an accuracy of 60.58%. The last experiment using the pixels per cell parameter of 8.8 and cells per block of 2.2 got the best accuracy of 61.43%.

This is an open access article under the CC BY-SA license.



Keywords:

KNN

HOG

skin cancer

Penulis Korespondensi:

Sarwati Rahayu

Fakultas Ilmu Komputer

Universitas Mercu Buana, Indonesia

Email: sarwati@mercubuana.ac.id

1. PENDAHULUAN

Sekitar lebih dari satu juta kasus kanker kulit terjadi pada tahun 2018 secara global. Kanker kulit adalah salah satu penyakit dengan pertumbuhan tercepat di dunia. Kanker kulit terjadi terutama karena paparan radiasi ultraviolet yang dipancarkan dari matahari [1]–[6]. Selain itu, jumlah pasien kanker meningkat akibat perilaku dalam merokok, adanya perubahan lingkungan, dan pengaruh faktor lain seperti radiasi, virus dan alkohol. Jenis kanker yang paling umum dan berbahaya adalah kanker kulit. Kanker kulit bisa dalam bentuk pembengkakan sel-sel kulit yang tidak biasa. Kanker kulit menyebar ke seluruh dunia dan merupakan penyakit berbahaya [7], [8].

Menurut WHO, kasus melanoma yang didiagnosis setiap tahun telah meningkat sebesar 53%, dan tingkat kematiannya akan terus meningkat. Kesalahan diagnosis dini menyebabkan tingkat kesembuhan dan indeks kelangsungan hidup kurang dari 14%. Namun jika deteksi kanker kulit dapat dilakukan sejak dini, maka indeks kelangsungan hidup dapat meningkat sekitar 97%. Laporan tentang masalah kanker kulit terus berkembang dengan jenis kanker yang paling umum adalah non-melanosit seperti karsinoma sel basal (*basal cell carcinoma*) dan karsinoma sel skuamosa (*squamous cell carcinoma*) [9], [10].

Sementara itu, kanker kulit *non-melanocytic* adalah bentuk dasar yang dapat ditemukan, seperti karsinoma sel basal dan karsinoma sel skuamosa [6]. Untuk meningkatkan tingkat penyembuhan, diagnosis dini adalah hal utama karena jika kanker ini semakin lama dideteksi maka penyakit telah menembus cukup jauh ke dalam kulit maka dapat sulit disembuhkan dan menyebabkan kematian [11]. Diagnosis kanker kulit dini sangat membantu untuk pengobatan kanker, meskipun tetap ada kemungkinan keadaan tersebut tidak dapat disembuhkan ketika penyakit tersebut telah berkembang di luar kulit [12].

Diagnosis juga tergantung pada etnis, paparan sinar matahari, kebiasaan sosial, dan riwayat medis pasien. Prosedur medis untuk diagnosis kanker kulit memakan waktu dan sangat sulit karena pemrosesan lesi biopsi (*biopsied lesions*). Dalam perkembangan teknologi kedokteran, teknologi kecerdasan artifisial memberikan kontribusi yang signifikan, terutama dalam pencitraan medis untuk menganalisis secara otomatis citra biopsied lesions.

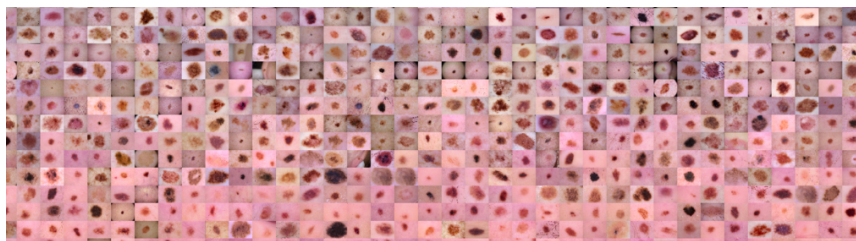
Penelitian mengenai analisis dataset *multiclass* kanker kulit untuk deteksi kanker kulit telah dilakukan oleh beberapa peneliti [8], [13]–[15]. Penelitian oleh Huang et al., (2021) membahas mengenai deteksi kanker kulit dengan menggunakan algoritma *deep learning*. Algoritma ini digunakan untuk mendukung pengembangan aplikasi medis berbasis *deep learning*. Pada penelitian ini dataset yang digunakan adalah HAM10000 dermoscopy. Hasil dari penelitian ini mendapatkan akurasi sebesar 85,8% [13].

Penelitian oleh Popescu et al. (2022) mengusulkan algoritma pembelajaran mendalam berbasis *convolutional neural network* (CNN) untuk menganalisis dataset kanker kulit HAM10000, yang dapat membedakan lesi kulit, termasuk melanoma. Penelitian ini menganalisis berbagai model CNN untuk bagaimana hasil *weight matrix*, dan elemen-elemennya pada *neural network* [14]. Selanjutnya penelitian oleh Srinivasu et al. (2021) menggunakan algoritma *mobileNet* dan *long short-term memory* untuk menganalisis dataset kulit HAM10000. Penelitian ini mendapatkan akurasi sebesar 85% [15], [16].

Mengingat ketersediaan sumber daya yang terbatas, deteksi dini kanker kulit sangat penting. Diagnosis yang akurat dan kelayakan deteksi sangat penting secara umum untuk kebijakan pencegahan kanker kulit. Deteksi kanker kulit pada fase awal merupakan tantangan bahkan bagi dokter kulit. Berdasarkan latar belakang diatas, penelitian ini akan menggunakan teknologi model kecerdasan artifisial yaitu analisis performa metode klasifikasi pada dataset *multiclass* kanker kulit menggunakan *k-nearest neighbor* (KNN) dan *histogram of oriented gradients* (HOG).

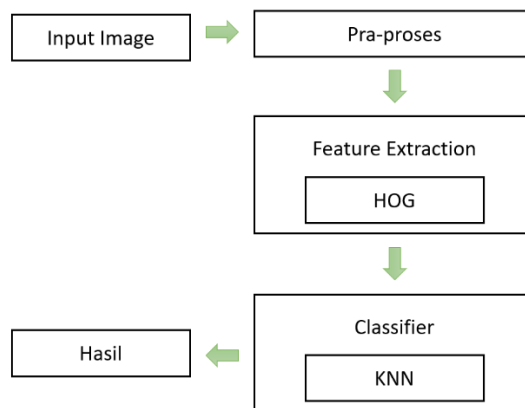
2. METODE PENELITIAN

Penelitian yang diusulkan bekerja pada 7 kelas dari kanker kulit antara lain *actinic keratoses and intraepithelial carcinoma/Bowen's disease* (akiec), *basal cell carcinoma* (bcc), *benign keratosis-like lesions* (bkl), *dermatofibroma* (df), *melanoma* (mel), *melanocytic nevi* (nv), dan *vascular lesions* (vasc). Dataset diambil secara publik dengan nama dataset Skin Cancer MNIST: HAM10000 dataset berjumlah 10,015 data, seperti yang terlihat pada **Gambar 1**.



Gambar 1 Contoh dataset

Metodologi penelitian yang diusulkan untuk deteksi kanker kulit dibagi menjadi tiga tahapan. Penelitian ini akan menggunakan teknologi model kecerdasan artifisial yaitu analisis performa metode klasifikasi pada dataset *multiclass* kanker kulit menggunakan *k-nearest neighbor* (KNN) dan *histogram of oriented gradients* (HOG). Ketiga fase ini dijelaskan di bawah ini secara rinci pada **Gambar 2**.



Gambar 2 Tahap Penelitian

Pertama, gambar input diteruskan ke tahap prapemrosesan kemudian pada fase berikutnya fitur dari citra kanker kulit diekstraksi. Teknik ekstraksi fitur *histogram of oriented gradients* (HOG) menggunakan distribusi arah gradien sebagai fitur. Fitur penting dari citra kanker kulit seperti gradien berorientasi diekstraksi menggunakan HOG [17]. Gradien citra kanker kulit bermanfaat karena besarnya gradien besar di sekitar tepi dan sudut karena hanya ada intensitas yang berubah secara tiba-tiba. Dalam menganalisis citra kanker kulit, informasi tepi dan sudut dianggap karena diperlukan untuk menentukan bentuk objek secara lengkap.

Selanjutnya, algoritma *k-nearest neighbor* (KNN) digunakan untuk pengukur jarak dalam mengklasifikasikan citra kanker kulit. Dalam tahap pelatihan, fitur vektor citra kanker kulit dibagi menjadi beberapa wilayah dan kemudian sesuai dengan kesamaan yang ada di antara mereka digabungkan ke wilayah tertentu. Untuk perhitungan kesamaan ukuran jarak digunakan pada citra kanker kulit, pada dasarnya dapat bekerja dengan baik jika pada kumpulan data kecil. Algoritma KNN menempatkan vektor pelatihan terdekat k terlepas dari label. Ketika citra kanker kulit baru X_i diberikan untuk algoritma klasifikasi yang ditempatkan untuk k tetangga terdekat membentuk titik citra kanker kulit baru menggunakan fungsi jarak k yang telah ditentukan [18].

3. HASIL DAN ANALISIS

Eksperimen pertama menggunakan metode *histogram of oriented gradients* (HOG) dengan parameter *pixels per cell* yaitu 8,8 dan *cells per block* yaitu 2,2. Dengan menggunakan metode HOG, citra kanker kulit dibagi menjadi *patch* 8×8 piksel yang tidak tumpang tindih yang disebut sel, di mana gradien dihitung pada setiap piksel. Untuk setiap sel, normalisasi dilakukan di seluruh blok 2×2 sel yang menghasilkan vektor fitur citra. Hasil dari eksperimen pertama dapat dilihat pada **Gambar 3**.

	precision	recall	f1-score	support	Confusion Matrix							
akiec	0.20	0.01	0.02	86	[1	0	14	5	0	66	0]
bcc	0.30	0.05	0.08	142	[0	7	12	10	0	112	1]
bkl	0.21	0.18	0.19	278	[1	1	50	15	6	204	1]
df	0.05	0.22	0.08	23	[0	1	4	5	0	13	0]
mel	0.23	0.03	0.05	285	[0	3	40	9	8	224	1]
nv	0.69	0.88	0.77	1649	[3	11	116	49	20	1446	4]
vasc	0.00	0.00	0.00	41	[0	0	3	4	1	33	0]]

(a)

(b)

Gambar 3 Hasil performa per kelas (a) dan *confusion matrix* (b) dari eksperimen pertama

Metode deteksi objek biasanya mengubah citra dari ruang piksel ke ruang fitur dimensional. Selanjutnya, algoritma klasifikasi digunakan untuk menentukan klasifikasi kelas kanker kulit. Metode *histogram of oriented gradients* (HOG) merupakan salah satu metode ekstraksi fitur yang digunakan secara luas untuk deteksi objek. Fitur HOG memberikan *trade-off* yang baik antara akurasi dan kompleksitas deteksi dibandingkan dengan metode lain.

Eksperimen kedua menggunakan metode *histogram of oriented gradients* (HOG) dengan parameter *pixels per cell* yaitu 16,16 dan *cells per block* yaitu 2,2. Dengan menggunakan metode HOG, citra kanker kulit dibagi menjadi *patch* 16×16 piksel yang tidak tumpang tindih yang disebut sel, di mana gradien dihitung pada setiap piksel. Untuk setiap sel, normalisasi dilakukan di seluruh blok 2×2 sel yang menghasilkan vektor fitur citra. Hasil dari eksperimen pertama dapat dilihat pada **Gambar 4**.

	precision	recall	f1-score	support	Confusion Matrix
akiec	0.20	0.01	0.02	86	<pre> [[1 0 10 2 2 70 1] [0 8 12 6 0 114 2] [1 1 51 12 6 207 0] [0 1 4 3 0 15 0] [0 4 37 4 11 229 0] [3 12 118 51 17 1443 5] [0 0 3 5 2 31 0]] </pre>
bcc	0.31	0.06	0.10	142	
bkl	0.22	0.18	0.20	278	
df	0.04	0.13	0.06	23	
mel	0.29	0.04	0.07	285	
nv	0.68	0.88	0.77	1649	
vasc	0.00	0.00	0.00	41	

(a)

(b)

Gambar 4 Hasil performa per kelas (a) dan confusion matrix (b) dari eksperimen kedua

Eksperimen ketiga menggunakan metode *histogram of oriented gradients* (HOG) dengan parameter *pixels per cell* yaitu 32,32 dan *cells per block* yaitu 2,2. Dengan menggunakan metode HOG, citra kanker kulit dibagi menjadi *patch* 16x16 piksel yang tidak tumpang tindih yang disebut sel, di mana gradien dihitung pada setiap piksel. Untuk setiap sel, normalisasi dilakukan di seluruh blok 2x2 sel yang menghasilkan vektor fitur citra. Hasil dari eksperimen pertama dapat dilihat pada Gambar 5.

	precision	recall	f1-score	support	Confusion Matrix
akiec	0.27	0.05	0.08	86	<pre> [[4 3 9 4 1 65 0] [2 11 11 5 1 112 0] [4 2 50 16 4 202 0] [0 1 3 4 0 15 0] [1 1 37 7 14 225 0] [4 12 113 40 25 1452 3] [0 0 3 4 3 30 1]] </pre>
bcc	0.37	0.08	0.13	142	
bkl	0.22	0.18	0.20	278	
df	0.05	0.17	0.08	23	
mel	0.29	0.05	0.08	285	
nv	0.69	0.88	0.77	1649	
vasc	0.25	0.02	0.04	41	

(a)

(b)

Gambar 5 Hasil performa per kelas (a) dan confusion matrix (b) dari eksperimen ketiga

Parameter orientasi pada HOG adalah jumlah buckets yang akan dibuat. Jika akan membuat matriks 9 x 1, maka dapat diatur nilai orientasi ke 9. Parameter *pixels per cell* mendefinisikan ukuran sel untuk membuat histogram. Sebagai contoh, penelitian akan menggunakan 8 x 8 sel, sehingga nilai *pixels per cell* dapat didefinisikan nilai yang sama. Selanjutnya, parameter *cells per block* hyperparameter yang merupakan ukuran blok untuk menormalkan histogram. Parameter ini mendefinisikan sel per blok, bukan jumlah piksel. Hasil eksperimen secara keseluruhan dapat dilihat pada Tabel 1.

Tabel 1 Hasil eksperimen HOG dan KNN

HOG					Klasifier	Akurasi
orientation	pixels per cell	cells per block	visualize	multichannel		
9	8, 8	2, 2	True	True	KNN	60,58%
9	16, 16	2, 2	True	True	KNN	60,58%
9	32, 32	2, 2	True	True	KNN	61,43%

4. KESIMPULAN

Mengingat ketersediaan sumber daya yang terbatas, deteksi dini kanker kulit sangat penting. Diagnosis yang akurat dan kelayakan deteksi sangat penting secara umum untuk kebijakan pencegahan kanker kulit. Deteksi kanker kulit pada fase awal merupakan tantangan bahkan bagi dokter kulit. Penelitian ini bertujuan untuk menanalisis performa metode klasifikasi pada dataset multiclass kanker kulit menggunakan *k-nearest neighbor* (KNN) dan *histogram of oriented gradients* (HOG). Dataset diambil secara publik dengan nama dataset Skin Cancer MNIST: HAM10000 dataset berjumlah 10,015 data. Eksperimen pertama menggunakan dengan parameter *pixels per cell* yaitu 8,8 dan *cells per block* yaitu 2,2 mendapatkan akurasi 60,58%. Eksperimen kedua menggunakan dengan parameter *pixels per cell* yaitu 8,8 dan *cells per block* yaitu 2,2 mendapatkan akurasi 60,58%. Eksperimen terakhir menggunakan dengan parameter *pixels per cell* yaitu 8,8 dan *cells per block* yaitu 2,2 mendapatkan akurasi paling baik yaitu 61,43%.

UCAPAN TERIMA KASIH

Penelitian ini merupakan luaran penelitian skema Penelitian Internal Kerjasama Dalam Negeri (KDN), Universitas Mercu Buana.

REFERENSI

- [1] R. Garg, S. Maheshwari, and A. Shukla, "Decision support system for detection and classification of skin cancer using CNN," in *Innovations in Computational Intelligence and Computer Vision*, Springer, 2021, pp. 578–586.
- [2] V. Ayumi, "Application of Machine Learning for SARS-CoV-2 Outbreak," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 7, no. 5, 2020.
- [3] V. Ayumi, "Studi Pendahuluan: Pengembangan Aplikasi m-BCARE Untuk Pasien Penderita Kanker Payudara," *JUSIBI (Jurnal Sist. Inf. dan E-Bisnis)*, vol. 3, no. 1, pp. 26–33, 2021.
- [4] D. Ramayanti *et al.*, "Tuberculosis Ontology Generation and Enrichment Based Text Mining," in *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2020, pp. 429–434.
- [5] D. I. Sensuse, P. Kareen, H. Noprisson, and M. O. Pratama, "Success factors for health information system development," in *2017 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2017, pp. 162–167.
- [6] I. Nurhaida, V. Ayumi, H. Noprisson, A. Ratnasari, M. Utami, and E. D. Putra, "Web Development Using WISDM and RAD," in *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2020, pp. 57–61.
- [7] A. M. De Marzo *et al.*, "Inflammation in prostate carcinogenesis," *Nat. Rev. Cancer*, vol. 7, no. 4, pp. 256–269, 2007.
- [8] T. M. Alam *et al.*, "An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset," *Diagnostics*, vol. 12, no. 9, p. 2115, Aug. 2022.
- [9] U. Leiter, T. Eigentler, and C. Garbe, "Epidemiology of skin cancer," *Sunlight, Vitam. D Ski. cancer*, pp. 120–140, 2014.
- [10] K. N. Syrigos, I. Tzannou, N. Katirtzoglou, and E. Georgiou, "Skin cancer in the elderly," *In Vivo (Brooklyn)*, vol. 19, no. 3, pp. 643–652, 2005.
- [11] S. Jain, U. Singhanian, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep Learning-Based Transfer Learning for Classification of Skin Cancer," *Sensors*, vol. 21, no. 23, p. 8142, 2021.
- [12] D. N. H. Thanh, V. B. Prasath, L. M. Hieu, and N. N. Hien, "Melanoma skin cancer detection method based on adaptive principal curvature, colour normalisation and feature extraction with the ABCD rule," *J. Digit. Imaging*, vol. 33, no. 3, pp. 574–585, 2020.
- [13] H. Huang, B. W. Hsu, C. Lee, and V. S. Tseng, "Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers," *J. Dermatol.*, vol. 48, no. 3, pp. 310–316, 2021.
- [14] D. Popescu, M. El-Khatib, and L. Ichim, "Skin Lesion Classification Using Collective Intelligence of Multiple Neural Networks," *Sensors*, vol. 22, no. 12, p. 4399, 2022.
- [15] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, 2021.
- [16] I. Nurhaida *et al.*, "Implementation of Deep Learning Predictor (LSTM) Algorithm for Human Mobility Prediction," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 18, p. 132, Nov. 2020.
- [17] S. P. Deore and A. Pravin, "Histogram of Oriented Gradients Based Off-Line Handwritten Devanagari Characters Recognition Using SVM, K-NN and NN Classifiers.," *Rev. d'Intelligence Artif.*, vol. 33, no. 6, pp. 441–446, 2019.
- [18] S. Saeed, A. Abdullah, N. Z. Jhanjhi, M. Naqvi, and A. Nayyar, "New techniques for efficiently k-NN algorithm for brain tumor detection," *Multimed. Tools Appl.*, vol. 81, no. 13, pp. 18595–18616, 2022.