

IDENTIFIKASI PENYAKIT MONKEYPOX MENGGUNAKAN ALGORITMA RANDOM FOREST DENGAN EKSTRAKSI FITUR GLCM

Muhammad Azka Zaki¹⁾

1) Program Studi Informatika Universitas Pembangunan Nasional "Veteran" Jawa Timur
E-mail: muhammadazkazaki@gmail.com

Abstrak

Monkeypox merupakan penyakit menular yang dapat dikenali melalui citra kulit penderita. Diperlukan metode diagnosis cepat dan akurat dalam mengidentifikasi penyakit Monkeypox. Penelitian ini bertujuan mengidentifikasi citra Monkeypox menggunakan metode ekstraksi fitur Gray Level Co-occurrence Matrix (GLCM) yang kemudian diklasifikasikan dengan algoritma Random Forest. Dataset terdiri dari 770 citra asli, yang kemudian diperluas menjadi 5860 citra melalui augmentasi transformasi geometri. Hasil pengujian menunjukkan bahwa metode ekstraksi fitur GLCM dengan klasifikasi Random Forest mampu mencapai akurasi sebesar 79%, menunjukkan performa cukup baik. Temuan ini menunjukkan potensi metode tersebut sebagai pendekatan komputasi sederhana dalam mendeteksi Monkeypox dan dapat dikembangkan dengan pendekatan kecerdasan buatan lainnya.

Kata kunci: Monkeypox, Medical Image Classification, GLCM, Random Forest, Feature Extraction

Pendahuluan

Monkeypox adalah penyakit kulit menular yang disebabkan oleh virus yang ditularkan dari binatang ke manusia. Penyakit ini memiliki dua fase. Fase pertama dikenal sebagai fase prodromal atau awal, berlangsung selama 1 hingga 3 hari dan ditandai dengan sakit kepala hebat, demam, dan pembengkakan kelenjar getah bening di leher, ketiak, atau selangkangan. Fase kedua dikenal sebagai fase erupsi. Pada fase ini, ruam pada kulit muncul dimulai di wajah dan secara bertahap menyebar ke seluruh tubuh. Mula-mula, ruam terbentuk menjadi bintik kecil yang menyerupai cacar, yang kemudian berkembang menjadi lepuhan kecil yang berair. Setelah bernanah dan mengeras, mereka kemudian rontok [1].

Penggunaan *Artificial Intelligent* sangat efektif, efisien, dan efektif dalam bidang Kesehatan [2]. Tidak terkecuali dalam hal penemuan penyakit kulit. *Gray Level Co-occurrence Matrix* (GLCM) adalah teknik yang cukup umum digunakan. Penelitian berjudul "Diagnosis kanker kulit menggunakan teknik pembelajaran mesin" menemukan bahwa penelitian ini mendiagnosis penyakit berdasarkan citra kulit menggunakan ekstraksi fitur GLCM yang dibandingkan dengan berbagai metode ekstraksi fitur lainnya. Penelitian ini mencapai hasil tertinggi dengan menggunakan metode Support Vector Machine yang dikombinasikan Random Forest, yang mendapat nilai akurasi tertinggi yaitu 89.31% untuk ekstraksi fitur GLCM [3].

Studi Pustaka

Monkeypox

Monkeypox adalah penyakit zoonosis yang disebabkan oleh genus Orthopoxvirus dari keluarga Poxviridae [4]. Virus ini dapat menyebar dari sumber zoonosis utama ke manusia. Kontak langsung, kontak dengan hewan, dan konsumsi daging yang kurang matang adalah beberapa cara penyakit ini dapat menyebar. Demam, nyeri kepala dan nyeri otot hingga punggung, serta ruam pada kulit adalah gejala penyakit monkeypox. Penyakit ini biasanya memiliki gejala yang berlangsung selama dua hingga empat minggu. Cacar monyet memiliki resiko kematian yang tinggi dan sulit untuk dibedakan dari jenis cacar lainnya.

Lesi cacar monyet memiliki diameter sekitar 0,5 hingga 1 cm, dan pola sentrifugal menunjukkan bahwa jumlah lesi yang tersebar dapat mencapai ribuan. Pola ini dimulai di wajah, lalu di telapak tangan dan kaki, serta membran mukosa mulut, konjungtiva, kornea, dan bahkan area kelamin.

Gray Level Co-occurrence Matrix

Tekstur dan bentuk gambar di bidang pengenalan pola dan visi komputer diukur dengan GLCM (*Gray Level Co-occurrence Matrix*). Tujuan dari GLCM adalah untuk menemukan *co-occurrence* dari level keabuan dua titik dalam gambar[5]. Matriks ini bekerja dengan kemungkinan bahwa dua titik tertentu muncul pada tingkat keabuan dalam jarak dan orientasi sudut tertentu [6]. Jarak antar titik dalam matriks dapat diatur tergantung kebutuhan dan kesesuaian pola citra. Dengan fitur 0° , 45° , 90° , dan 135° , GLCM dibagi menjadi empat arah. GLCM juga memiliki *properties* atau fitur yang diambil dari citra. *Properties* yang digunakan dalam penelitian ini meliputi:

- a. *Contrast*: Jumlah keberagaman intensitas level keabuan dalam gambar.
- b. *Energy*: Ukuran konsentrasi pasangan intensitas dalam gambar.
- c. *Homogeneity*: Nilai keseragaman atau keabuan sejenis dalam gambar.
- d. *Correlation*: Nilai yang menunjukkan adanya struktur linier dalam gambar.

Random Forest

Random Forest adalah algoritma klasifikasi dengan konsep banyak *decision tree* dan termasuk pada metode *ensemble learning* [7]. Algoritma ini merupakan pengembangan dari algoritma *decision tree*, dimana banyak pohon Keputusan dibangun sekaligus dan dilatih menggunakan subset data yang acak.

Confusion Matrix

Confusion Matrix menjelaskan bagaimana sebuah model melakukan klasifikasi dan kesalahan. *Tools* ini bekerja dengan cara mengkomparasi nilai aktual dari data dengan nilai hasil prediksi model [8]. Beberapa *performance metrics* dihitung berdasarkan hasil dari *confusion matrix*, meliputi:

- a. *Akurasi*

Nilai persentase dari data yang terprediksi dengan benar terhadap keseluruhan data, dirumuskan sebagai berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- b. *Presisi*

Nilai persentase dari perbandingan jumlah data yang diklasifikasikan ke kelas positif dengan jumlah keseluruhan data yang terkласifikasi ke kelas positif, dirumuskan sebagai berikut.

$$Precision = \frac{TP}{FP+TP} \quad (2)$$

- c. *Recall*

Nilai persentase dari perbandingan jumlah data yang diklasifikasi secara benar ke dalam kelas positif dengan seluruh data yang terklasifikasi ke kelas positif, dirumuskan sebagai berikut.

$$Recall = \frac{TP}{FN+TP} \quad (3)$$

- d. *F1-score*

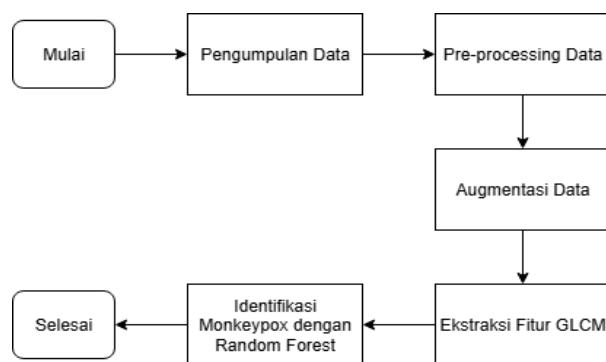
Nilai kombinasi dari presisi dan *recall* untuk mendapat Gambaran keseimbangan dari presisi dan *recall*, dirumuskan sebagai berikut.

$$F1 score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Metodologi Penelitian

Alur penelitian dimulai dengan pengumpulan data. Pada penelitian ini, dataset yang digunakan merupakan dataset bernama “*Monkeypox Skin Images Dataset (MSID)*” yang diambil dari situs *kaggle* [9]. Dataset ini memiliki 770 gambar yang terbagi dalam empat kelas, yaitu kelas *chickenpox* berisi 107 citra, kelas *measles* berisi 91 citra, kelas *monkeypox* berisi 279 citra, dan kelas

normal berisi 293 citra. Untuk menyesuaikan format gambar dengan metode ekstraksi fitur GLCM, dilakukan *preprocessing* berupa konversi format gambar menjadi *grayscale*. Kemudian, untuk menyeimbangkan jumlah data yang tidak seimbang pada tiap-tiap kelasnya, dilakukan augmentasi dengan oversampling. Augmentasi yang dilakukan adalah transformasi geometris. Jumlah data setelah augmentasi dengan oversampling adalah 5860 gambar yang terbagi dalam 4 kelas sama rata, tiap kelas memiliki 1465 gambar. Dilanjutkan dengan tahap ekstraksi fitur dengan GLCM. Data yang sudah melalui tahap *preprocessing* dan augmentasi dengan *oversampling* akan menjadi input untuk diambil ciri fiturnya. Fitur-fitur yang diambil dari citra meliputi *contrast*, *energy*, *homogeneity*, dan *correlation*, dengan beberapa parameter lain yaitu sudut bernilai 0° , 45° , 90° , dan 135° serta *distance* atau jarak antar piksel bernilai 1 dan 2. Setelah fitur diekstrak dengan GLCM, maka hasil fitur tersebut akan diolah dengan algoritma *Random Forest*. Parameter *Random Forest* yang digunakan adalah jumlah pohon yang dibangun, yaitu bernilai 100 dan 300. Parameter-parameter yang disebutkan tadi dikombinasikan sedemikian rupa untuk diambil model terbaik berdasarkan beberapa skenario. Gambar di bawah merupakan rangkaian alur atau metodologi penelitian.



Gambar 1. Metodologi penelitian

Hasil dan Pembahasan

Dilakukan 8 skenario berbeda untuk menemukan model yang memiliki nilai akurasi terbaik. Skenario disusun berdasarkan persentasi pembagian data uji - data latih bernilai 70:30 dan 80:20, *distance* piksel GLCM bernilai 1 dan 2, serta jumlah pohon yang dibentuk pada *Random Forest* 100:300, sehingga didapat tabel skenario seperti di bawah ini.

Tabel 1. Skenario Pengujian

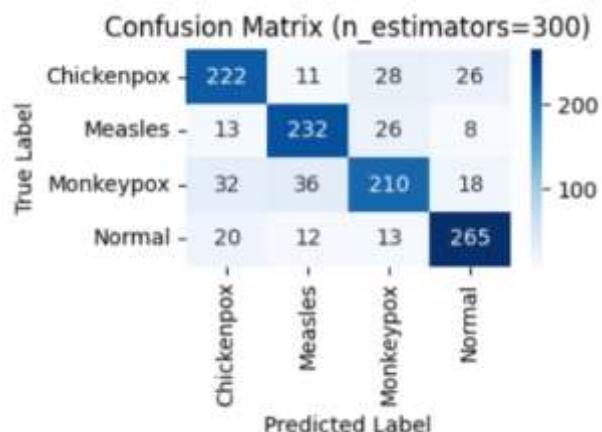
No.	Data Latih (%)	Data Uji (%)	<i>Distance</i> GLCM	Jumlah Pohon RF
1	70	30	1	100
2				300
3			2	100
4				300
5	80	20	1	100
6				300
7			2	100
8				300

Dilakukan evaluasi model yang sudah dijalankan berdasarkan skenario tabel 1, menggunakan matriks konfusi dengan nilai akurasi, presisi, *recall*, dan *F1-score*. Kemudian diambil model dengan nilai akurasi tertinggi.

Tabel 2. Hasil Evaluasi Model

No. Skenario	Akurasi	Presisi	Recall	F1-Score
1	75	75	75	75
2	76	76	76	76
3	77	77	77	77
4	77	77	77	77
5	77	77	77	77
6	77	77	77	77
7	78	78	78	78
8	79	79	79	79

Berdasarkan tabel 2, diketahui bahwa model dengan skenario ke 8 memiliki nilai akurasi tertinggi ketika dibandingkan dengan model skenario yang lain. Akurasi yang didapat bermilai 79%, dengan nilai presisi 79%, nilai *recall* 79%, dan nilai F1-score 79%. Model skenario 8 dijalankan dengan kombinasi parameter perbandingan data latih dan data uji 80:20, *distance* GLCM bermilai 2, serta pohon *Random Forest* yang dibentuk berjumlah 300.

Gambar 2. *Confusion matrix* model skenario 8

Confusion Matrix dari model skenario ke 8 ditampilkan pada gambar 2 di atas. Terlihat bahwa model tersebut mampu memprediksi 929 data dengan benar, yang terdiri dari 222 data kelas chickenpox, 232 data kelas measles, 210 data kelas monkeypox, dan 265 data kelas normal. Sedangkan data yang terprediksi salah sebanyak 243.

Kesimpulan

Berdasarkan hasil pengujian model, diketahui bahwa metode ekstraksi fitur GLCM dengan klasifikasi *Random Forest* mendapat akurasi tertinggi senilai 79%. Ini menunjukkan bahwa kombinasi dua metode tersebut cocok untuk digunakan, terutama pada penyakit Monkeypox. Meskipun dua metode tersebut cocok untuk dikombinasikan, tetapi masih ada ruang untuk optimalisasi menggunakan metode kecerdasan buatan yang lain.

Daftar Pustaka

- [1] Dinas Kesehatan Provinsi Gorontalo, "WHO Tetapkan Monkeypox Darurat Kesehatan Global, Kadinkes dr. Yana : Tingkatkan Upaya Promotif dan Preventif." Accessed: Jan. 24, 2025. [Online]. Available: <https://dinkes.gorontaloprov.go.id/who-tetapkan-monkeypox-darurat-kesehatan-global-kadinkes-dr-yana-tingkatkan-upaya-promotif-dan-preventif/>
- [2] M. A. Thaariq¹, M. Dimas, M. Baskara², R. A. Chaniago³, D. Christin⁴, and I. Ernawati, "SYSTEMATIC LITERATURE REVIEW: ANALISIS PENERAPAN KECERDASAN BUATAN DALAM BIDANG KESEHATAN," 2024.
- [3] A. Murugan, S. A. H. Nair, A. A. P. Preethi, and K. P. S. Kumar, "Diagnosis of skin cancer using machine learning techniques," *Microprocess Microsyst*, vol. 81, Mar. 2021, doi: 10.1016/j.micpro.2020.103727.
- [4] L. Budiyarto, A. A. Sabil, and H. C. Putri, "INFEKSI CACAR MONYET (MONKEYPOX)." [Online]. Available: <http://jurnalmedikahutama.com>
- [5] Z. Y. Lamasigi and A. Bode, "Influence of gray level co-occurrence matrix for texture feature extraction on identification of batik motifs using k-nearest neighbor," *ILKOM Jurnal Ilmiah*, vol. 13, no. 3, pp. 322–333, Dec. 2021, doi: 10.33096/ilkom.v13i3.1025.322-333.
- [6] M. Muhathir, M. H. Santoso, and D. A. Larasati, "Wayang Image Classification Using SVM Method and GLCM Feature Extraction," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 4, no. 2, pp. 373–382, Jan. 2021, doi: 10.31289/jite.v4i2.4524.
- [7] N. Faoziatun Khusna *et al.*, "Implementasi Random Forest dalam Klasifikasi Kasus Stunting pada Balita dengan Hyperparameter Tuning Grid Search," *Seminar Nasional Sains Data*, vol. 2024.
- [8] A. Zafar *et al.*, "A Comparison of Pooling Methods for Convolutional Neural Networks," *Applied Sciences*, vol. 12, no. 17, p. 8643, Aug. 2022, doi: 10.3390/app12178643.
- [9] D. Bala *et al.*, "MonkeyNet: A robust deep convolutional neural network for monkeypox disease detection and classification," *Neural Networks*, vol. 161, pp. 757–775, Apr. 2023, doi: 10.1016/j.neunet.2023.02.022.