



Identifikasi Citra Penyakit Monkeypox dengan Random Forest Serta Ekstraksi Fitur VGG19

Muhammad Azka Zaki ^{1*}, Eka Prakarsa Mandyartha ², dan Achmad Junaidi ³

¹ UPN "Veteran" Jawa Timur, e-mail : muhammadazkazaki@gmail.com

² UPN "Veteran" Jawa Timur, e-mail : eka_prakarsa.fik@upnjatim.ac.id

³ UPN "Veteran" Jawa Timur, e-mail : achmadjunaidi.ifi@upnjatim.ac.id

* Corresponding Author : Muhammad Azka Zaki

Abstract: Monkeypox is an infectious disease that can be recognized through images of the patient's skin lesions. A fast and accurate diagnosis method is required to identify Monkeypox. This research aims to identify Monkeypox imagery using the VGG19 feature extraction method, which is then classified using the Random Forest algorithm. The dataset consists of 770 original images, which were expanded to 5,860 images through geometric transformation augmentation. The test results show that the VGG19 feature extraction method with Random Forest classification achieved an accuracy of 95.1%, indicating good performance. This finding suggests the potential of this method as a machine learning approach for detecting Monkeypox and can be further developed with other artificial intelligence approaches.

Keywords: Monkeypox; Feature Extraction; VGG19; Classification; Random Forest

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 VGG19 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 VGG19 dengan klasifikasi Random Forest mampu mencapai nilai akurasi sebesar 95.1%, menunjukkan performa cukup baik. Temuan ini menunjukkan potensi metode tersebut sebagai pendekatan machine learning dalam mendeteksi Monkeypox dan dapat dikembangkan dengan pendekatan kecerdasan buatan lainnya.

Kata kunci: Monkeypox; Ekstraksi Fitur; VGG19; Klasifikasi; Random Forest

Naskah Masuk: 29 Desember 2025

Revisi: 6 Januari 2026

Diterima: 14 Maret 2026

Terbit: 17 Maret 2026

Ver. Skrg.: 17 Maret 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. 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. Penelitian berjudul "Monkeypox Detection Using CNN with Transfer Learning" berfokus pada membandingkan beberapa model transfer learning yang terdiri dari EfficientNetV2s, MobileNetV3, VGG19, ResNet50, DenseNet, dan Xception dalam mendeteksi penyakit monkeypox melalui citra lesi

dengan deep learning. Hasil dari penelitian ini didapatkan bahwa MobileNetV3-s mendapat skor tertinggi dengan akurasi senilai 96%. VGG19 juga mendapat skor akurasi yang cukup tinggi yaitu 92% [3].

Penelitian berjudul “Feature Extraction Using Hybrid Approach of VGG19 and GLCM for Optimized Brain Tumor Classification” bertujuan meningkatkan akurasi klasifikasi tumor otak. Metode yang digunakan dalam penelitian ini adalah Gray Level Co-Occurrence Matrix dan VGG19 yang digunakan untuk mengekstraksi fitur citra MRI otak. Pendekatan penelitian ini menggabungkan ekstraksi fitur Gray Level Co-Occurrence Matrix dan VGG19 kemudian mengklasifikasikannya dengan CNN. Hasil dari penelitian ini didapat bahwa kombinasi ekstraksi menggunakan metode GLCM dan VGG19 mendapat akurasi senilai 98% melebihi model AlexNet dan DenseNet 121 yang hanya mendapat akurasi di kisaran 50% [4].

Penelitian berjudul “Klasifikasi Monkeypox Menggunakan Ekstraksi Fitur GLCM dan Algoritma Random Forest” karya William Wijaya, Muhammad Rizky Pribadi, dan Eka Puji Widiyanto bertujuan untuk klasifikasi citra monkeypox dan non-monkeypox menggunakan ekstraksi fitur Gray Level Co-occurrence Matrix (GLCM) dan algoritma Random Forest. Akurasi tertinggi yang didapat pada penelitian ini senilai 77% yang menunjukkan ruang untuk peningkatan [5].

Penelitian ini berfokus untuk mengetahui bagaimana nilai akurasi dari penggunaan metode klasifikasi *Random Forest* dan VGG19 sebagai ekstraksi fitur dan membandingkannya dengan nilai akurasi dari penggunaan metode *Random Forest* dan GLCM pada penelitian sebelumnya.

2. Studi Pustaka

2.1. Monkeypox

Monkeypox adalah penyakit zoonosis yang disebabkan oleh genus Orthopoxvirus dari keluarga Poxviridae [6]. 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.

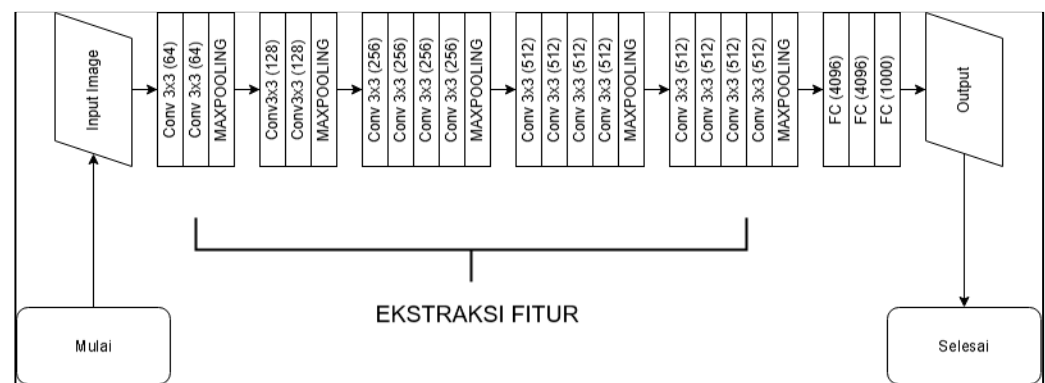
Lesi cacar monyet memiliki diameter sekitar 0,5 hingga 1 cm, dan pola sentrifugal menunjukkan bahwa jumlah lesi yang tersebar dapat mencapai ribuan. Pola ini dimulai di wajah, lalu di telapak tangan dan kaki, serta membran mukosa mulut, konjungtiva, kornea, dan bahkan area kelamin.

2.2 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.

2.3 VGG19

VGG19 merupakan salah satu arsitektur dari Convolutional Neural Network [8]. Arsitektur ini dikembangkan di Oxford University pada tahun 2014 oleh Visual Geometry Group. VGG19 terdiri dari 19 lapisan atau layer, dengan rincian 16 layer konvolusional dan 3 fully connected layer [9].



Gambar 1. Arsitektur VGG19

VGG19 adalah arsitektur yang pernah digunakan untuk melatih lebih dari 1 juta gambar yang didapatkan dari database ImageNet. Arsitektur ini memiliki kernel berukuran 3 x 3, dan setiap blok memiliki 5 blok convolutional layer dengan ukuran yang berbeda, masing-masing dipisahkan oleh lapisan max pooling [10]. Penggunaan kernel 3 x 3, menjadi keunggulan VGG19 dengan arsitektur lain. Kernel 3 x 3 mampu menghasilkan jaringan yang lebih kompleks dan dalam, sehingga dapat mempelajari fitur kompleks dengan lebih baik [11].

2.4 Confusion Matrix

Confusion Matrix menjelaskan bagaimana sebuah model melakukan klasifikasi dan kesalahan. Tools ini bekerja dengan cara mengomparasi nilai aktual dari data dengan nilai hasil prediksi model [12]. Beberapa performance metrics dihitung berdasarkan hasil dari confusion matrix, meliputi:

2.4.1. Accuracy

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

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

2.4.2. Precision

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

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

2.4.3. Recall

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

$$Recall = \frac{TP}{FN + TP} \quad (2)$$

2.4.4. F1-Score

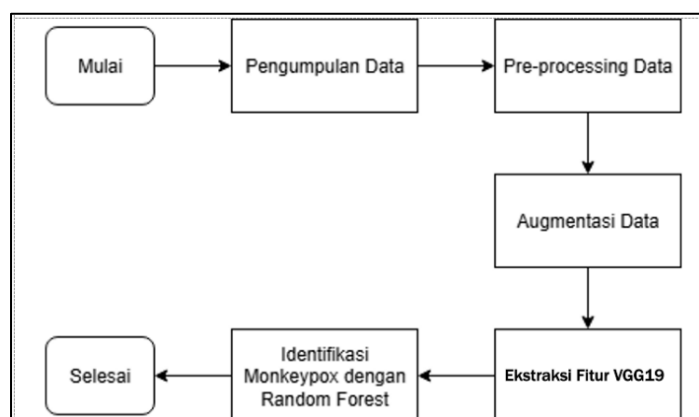
Nilai kombinasi dari presisi dan recall untuk mendapat Gambaran keseimbangan dari presisi dan recall, dirumuskan sebagai berikut.

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

3. Metodologi

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 VGG19, dilakukan preprocessing berupa konversi format gambar menjadi ukuran 224×224 . 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 VGG19. Data yang sudah melalui tahap preprocessing dan augmentasi dengan oversampling akan menjadi input untuk diambil ciri fiturnya. Fitur-fitur yang diambil dari citra meliputi merupakan layer FC1 ataupun FC2 yang terdiri dari 4096 fitur yang disimpan pada sebuah file berbentuk CSV. Setelah fitur diekstrak dengan VGG19, 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 tersebut digunakan atas dasar pertimbangan bahwa semakin banyak jumlah pohon maka semakin baik hasil yang diperoleh, namun akan membuat model semakin berat sehingga perlu ditentukan jumlah pohon yang ideal melalui beberapa skenario. 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 2. Flowchart Metodologi Penelitian

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, layer VGG19 yang diambil sebagai fitur yaitu layer Fully Connected 1 dan Fully Connected 2, serta jumlah pohon yang dibentuk pada Random Forest yaitu 100 dan 300, sehingga didapat tabel skenario seperti tabel 1.

Tabel 1. Skenario Pelatihan Model

No	Data Latih	Data Uji	Layer VGG19	Jumlah Pohon RF
1	70%	30%	FC1	100
2				300
3			FC2	100
4				300
5	80%	20%	FC1	100
6				300
7			FC2	100
8				300

Dilakukan evaluasi model menggunakan *confusion matrix* dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Kemudian diambil mode dengan nilai *accuracy* tertinggi.

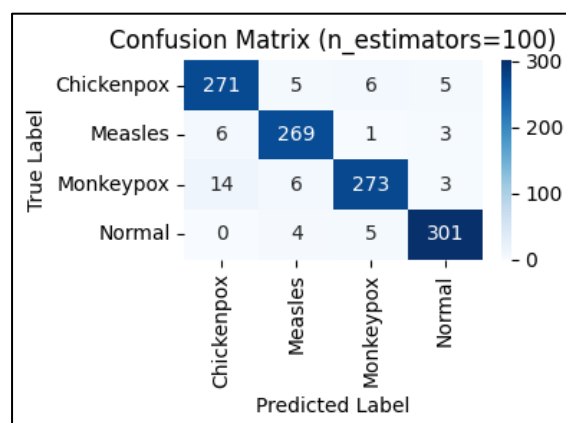
4. Hasil dan Pembahasan

Berdasarkan tabel 2, diketahui bahwa model pada skenario ke 5, 6, dan 8 memiliki nilai akurasi tertinggi ketika dibandingkan dengan model skenario yang lain. Nilai akurasi yang didapat bernilai 95.1%. Model skenario 5 dijalankan dengan kombinasi parameter perbandingan data latih dan data uji 80:20, layer VGG19 FC1, serta pohon Random Forest yang dibentuk berjumlah 100. Model skenario 6 dijalankan dengan kombinasi parameter perbandingan data latih dan data uji 80:20, layer VGG19 FC1, serta pohon Random Forest yang dibentuk berjumlah 300. Sedangkan model skenario 8 dijalankan dengan kombinasi parameter perbandingan data latih dan data uji 80:20, layer VGG19 FC2, serta pohon Random Forest yang dibentuk berjumlah 300.

Tabel 2. Hasil Pelatihan Model

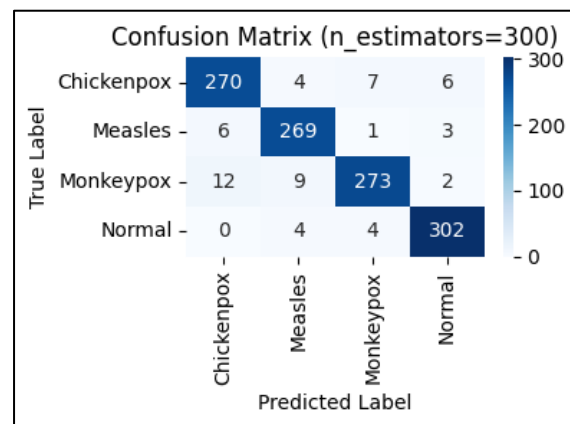
No	Akurasi	Presisi	Recall	F1-score
1	94.5%	95%	95%	95%
2	94.8%	95%	95%	95%
3	93.2%	93%	93%	93%
4	94.4%	94%	94%	94%
5	95.1%	95%	95%	95%
6	95.1%	95%	95%	95%
7	93.5%	94%	94%	94%
8	95.1%	95%	95%	95%

Confusion Matrix dari model skenario ke 5 ditampilkan pada gambar 3. Terlihat bahwa model tersebut mampu memprediksi 1114 data dengan benar, yang terdiri dari 271 data kelas chickenpox, 269 data kelas measles, 273 data kelas monkeypox, dan 301 data kelas normal. Sedangkan data yang terprediksi salah sebanyak 36.



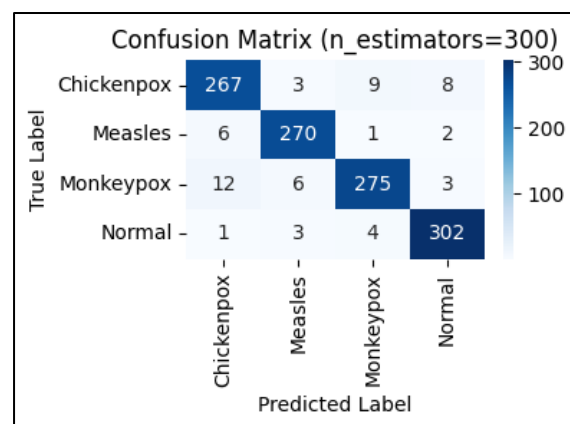
Gambar 3. Skenario Pelatihan Model 5

Confusion Matrix dari model skenario ke 6 ditampilkan pada gambar 4. Terlihat bahwa model tersebut mampu memprediksi 1114 data dengan benar, yang terdiri dari 270 data kelas chickenpox, 269 data kelas measles, 273 data kelas monkeypox, dan 302 data kelas normal. Sedangkan data yang terprediksi salah sebanyak 36.



Gambar 4. Skenario Pelatihan Model 6

Confusion Matrix dari model skenario ke 8 ditampilkan pada gambar 5. Terlihat bahwa model tersebut mampu memprediksi 1114 data dengan benar, yang terdiri dari 267 data kelas chickenpox, 270 data kelas measles, 275 data kelas monkeypox, dan 302 data kelas normal. Sedangkan data yang terprediksi salah sebanyak 36.



Gambar 5. Skenario Pelatihan Model 8

5. Kesimpulan

Berdasarkan hasil evaluasi model, diketahui bahwa metode ekstraksi fitur VGG19 dengan klasifikasi Random Forest mendapat nilai akurasi tertinggi senilai 95.1%, nilai ini lebih tinggi jika dibandingkan penelitian sebelumnya dimana penggunaan *Random Forest* dan GLCM mencapai 77%. Ini menunjukkan bahwa kombinasi metode klasifikasi *Random Forest* dan VGG19 sebagai ekstraksi fitur cocok untuk digunakan, terutama pada penyakit Monkeypox. Meskipun dua metode tersebut cocok untuk dikombinasikan, tetapi masih ada ruang untuk pengembangan lebih lanjut menggunakan metode kecerdasan buatan yang lain.

Kontribusi Penulis : Pada penelitian ini, Muhammad Azka Zaki berkontribusi sebagai mahasiswa sekaligus penulis utama. Eka Prakarsa Mandyartha berkontribusi sebagai dosen pembimbing 1. Achmad Junaidi berkontribusi sebagai dosen pembimbing 2.

Konflik Kepentingan : Para penulis menyatakan tidak ada konflik kepentingan.

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] M. Altun, H. Gürüler, O. Özkaraca, F. Khan, J. Khan, and Y. Lee, “Monkeypox Detection Using CNN with Transfer Learning,” *Sensors*, vol. 23, no. 4, p. 1783, Feb. 2023, doi: 10.3390/s23041783.
 - [4] M. SHARMA and S. BENIWAL, “Feature Extraction Using Hybrid Approach of VGG19 and GLCM For Optimized Brain Tumor Classification,” *International Journal of Computational and Experimental Science and Engineering*, vol. 10, no. 4, Dec. 2024, doi: 10.22399/ijcesen.714.
 - [5] W. Wijaya, M. R. Pribadi, and E. P. Widiyanto, “2 ND MDP STUDENT CONFERENCE (MSC) 2023 KLASIFIKASI MONKEYPOX MENGGUNAKAN EKSTRAKSI FITUR GLCM DAN ALGORITMA RANDOM FOREST”.
 - [6] L. Budiarto, A. A. Sabila, and H. C. Putri, “INFEKSI CACAR MONYET (MONKEYPOX).” [Online]. Available: <http://jurnalmedikahutama.com>
 - [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. Fendiawati and M. E. Al Rivan, “Klasifikasi American Sign Language Dengan Metode VGG-19,” *MDP Student Conference*, vol. 2, no. 1, pp. 192–197, Apr. 2023, doi: 10.35957/mdp-sc.v2i1.4466.
 - [9] F. Ramadhan and J. Hernadi, “Evaluasi Optimizer Adam dan RMSProp pada Arsitektur VGG-19 Klasifikasi Ekspresi Wajah Manusia,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1414–1426, Mar. 2025, doi: 10.29100/jupi.v10i2.6197.
 - [10] RAHMAD SYUHADA, “ANALISIS ARSITEKTUR DEEP LEARNING VGG UNTUK KLASIFIKASI JENIS JAMUR,” 2023.
 - [11] A. M. Syafar, “IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI TANAMAN RIMPANG SECARA VIRTUAL DARMATASIA,” vol. 8, no. 1, 2023.
 - [12] 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.