

Implementasi Algoritma Random Forest Untuk Klasifikasi Citra Medis Pada Kasus Penyakit Flu Singapura (HFMD)

Novita Christin Roulina Nainggolan¹, Indra Kelana Jaya², Harlen Gilbert Simanullang³

^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Sep 9, 2019

Revised, May 20, 2020

Accepted, Jun 11, 2020

Keywords:

HFMD, Flu Singapura, *Random Forest*, GLCM, klasifikasi citra medis.

ABSTRAK

Flu Singapura atau *Hand, Foot, and Mouth Disease* (HFMD) merupakan penyakit akibat infeksi virus yang menimbulkan lesi pada kulit di area tangan, kaki, dan mulut. Deteksi dini diperlukan untuk mencegah penyebaran dan mempercepat penanganan. Penelitian ini menerapkan algoritma *Random Forest* untuk mengklasifikasikan citra lesi HFMD guna meningkatkan akurasi deteksi penyakit. Dataset diperoleh dari *platform Kaggle*, kemudian melalui tahap preprocessing, ekstraksi fitur dengan *Gray Level Co-occurrence Matrix* (GLCM), dan klasifikasi menggunakan *Random Forest*. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Hasil menunjukkan akurasi sebesar 72,00%, dengan *precision* 75,56%, *recall* 77,27%, dan *F1-score* 76,40%. Berdasarkan hasil tersebut, algoritma *Random Forest* dinilai efektif dalam klasifikasi citra lesi HFMD dan dapat dikembangkan lebih lanjut untuk sistem deteksi dini berbasis kecerdasan buatan.

This is an open access article under the [CC BY-SA](#) license.



Penulis Koresponden:

Novita Christin Roulina Nainggolan,
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: nainggolannovita17@gmail.com

1. PENDAHULUAN

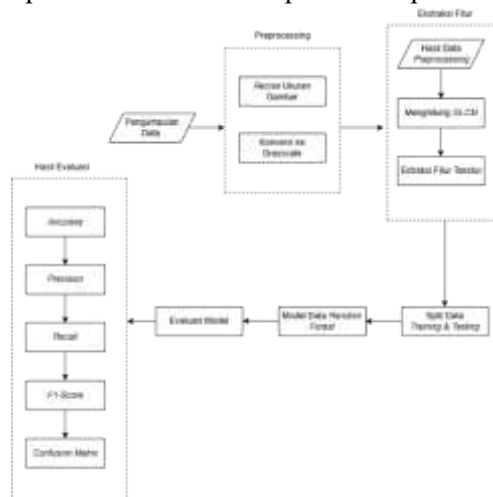
Hand, Foot, and Mouth Disease (HFMD) atau yang dikenal sebagai Flu Singapura adalah suatu penyakit menular yang disebabkan oleh virus *Coxsackievirus* dan *Enterovirus 71*. Penyakit ini ditandai dengan munculnya lesi pada kulit di bagian tangan, kaki, dan mulut, serta umumnya menyerang anak-anak. Meskipun sebagian besar kasus tergolong ringan, HFMD dapat menimbulkan komplikasi serius apabila tidak segera dideteksi dan ditangani secara tepat (*World Health Organization*). Oleh karena itu, deteksi dini menjadi langkah penting dalam upaya pencegahan dan pengendalian penyebaran penyakit ini. Secara global, HFMD banyak ditemukan di wilayah Asia dengan tingkat prevalensi yang tinggi. Setiap tahunnya, jutaan kasus dilaporkan di negara-negara seperti Tiongkok, Jepang, Malaysia, Singapura, Vietnam, dan Thailand (*World Health Organization*; Kementerian Kesehatan RI). Di Indonesia, data Kementerian Kesehatan menunjukkan bahwa kasus HFMD terus meningkat setiap tahun, menjadikannya sebagai permasalahan kesehatan masyarakat yang memerlukan perhatian lebih. Proses identifikasi manual HFMD melalui pengamatan lesi kulit membutuhkan keahlian khusus dan waktu yang cukup lama. Oleh sebab itu, teknologi berbasis pembelajaran mesin (*Machine Learning*) hadir sebagai alternatif yang menjanjikan. Pendekatan pengolahan citra medis (*medical image processing*) telah banyak diterapkan dalam mendeteksi berbagai kelainan kulit maupun penyakit menular lainnya. Berdasarkan uraian tersebut, penelitian ini dilakukan untuk mengevaluasi kinerja algoritma *Random Forest* dalam mengklasifikasikan citra lesi pada

penderita HFMD[1]. Penelitian ini diharapkan mampu memberikan kontribusi terhadap pengembangan teknologi deteksi dini berbasis kecerdasan buatan (AI) yang mampu mendukung proses diagnosis medis secara lebih efektif dan akurat.

2. METODE PENELITIAN

2.1. Desain Penelitian

Setiap langkah dalam proses disusun untuk memastikan tujuan penelitian dapat tercapai secara efektif. Desain ini juga memberikan rincian terkait dengan metode yang digunakan dalam setiap tahapan. Rancangan penelitian yang diterapkan dalam studi ini dapat dilihat pada Gambar 1.



Gambar 1. Rancangan Tahapan Penelitian

2.2. Pengumpulan Data

Tahap pengumpulan data dilakukan untuk memperoleh citra sebagai objek utama penelitian. Dataset diperoleh dari *platform Kaggle*, yang berisi kumpulan gambar dengan indikasi penyakit *Hand, Foot, and Mouth Disease* (HFMD). Citra yang masih berupa data mentah kemudian diproses melalui tahap pra-pemrosesan (*preprocessing*) agar sesuai dengan kebutuhan analisis dan pengolahan selanjutnya.

2.3. Pre-processing

Pengolahan citra digital adalah suatu proses yang dilakukan untuk memperbaiki atau meningkatkan mutu gambar sehingga lebih mudah dipahami dan dianalisis, baik secara visual oleh manusia maupun secara otomatis oleh sistem komputer[2]. Setiap citra digital terdiri atas piksel-piksel yang memiliki nilai intensitas atau warna tertentu, sehingga memungkinkan pengenalan elemen visual seperti bentuk, pola, dan tekstur[3]. Dalam konteks pembelajaran mesin, pengolahan citra meliputi beberapa tahapan utama, yaitu akuisisi citra, pra-pemrosesan (*pre-processing*), segmentasi, serta ekstraksi fitur yang menjadi dasar dalam proses pengenalan pola dan klasifikasi.

Tahap pre-processing dilakukan untuk mempersiapkan citra sebelum dianalisis lebih lanjut, dengan tujuan mengurangi noise serta menyeragamkan karakteristik visual. Pada penelitian ini, proses pre-processing mencakup beberapa langkah, yaitu:

1. Resizing, mengubah citra menjadi ukuran 128×128 piksel.
2. Grayscale, mengubah citra menjadi skala abu-abu.

2.4. Ekstraksi Fitur dengan *Gray Level Co-occurrence Matrix (GLCM)*

Tahap ekstraksi fitur bertujuan untuk mengambil informasi penting dari citra dalam bentuk yang lebih sederhana agar mudah dianalisis. Pada penelitian ini digunakan *metode Gray Level Co-occurrence Matrix (GLCM)*, yang berfungsi untuk mengidentifikasi tekstur citra berdasarkan hubungan spasial antar piksel[4]. GLCM menghitung seberapa sering pasangan piksel dengan tingkat keabuan tertentu muncul pada jarak dan arah tertentu, sehingga dapat merepresentasikan pola tekstur khas dari setiap citra[5]. Dari hasil perhitungan matriks GLCM, diperoleh sejumlah parameter tekstur yang mewakili ciri khas setiap citra, antara lain:

1. *Contrast* : Mengukur tingkat perbedaan intensitas antar piksel pada citra.
2. *Energy* : Menggambarkan tingkat keseragaman atau keteraturan pola tekstur.
3. *Homogeneity* : Menunjukkan seberapa dekat nilai intensitas antar piksel yang berdekatan.

4. *Correlation* : Menilai hubungan linear antara piksel-piksel pada posisi tertentu

Parameter-parameter tersebut selanjutnya dimanfaatkan sebagai fitur masukan (*input features*) dalam proses klasifikasi menggunakan algoritma *Random Forest*, dengan tujuan membedakan antara citra yang mengindikasikan penyakit *Hand, Foot, and Mouth Disease* (HFMD) dan citra normal.

2.5. Klasifikasi dengan *Random Forest*

Algoritma *Random Forest* termasuk dalam teknik ensemble learning yang sering dimanfaatkan pada bidang *machine learning*, terutama untuk menangani proses klasifikasi terhadap data berukuran besar dan memiliki tingkat kompleksitas tinggi. Prinsip kerjanya adalah membentuk sejumlah pohon keputusan (*decision tree*) secara acak dari data pelatihan, kemudian mengombinasikan hasil dari setiap pohon untuk memperoleh prediksi yang lebih konsisten serta memiliki tingkat akurasi yang tinggi[6]. Pada algoritma *Random Forest*, setiap pohon keputusan dibangun dengan memanfaatkan metode *bootstrap sampling*, yaitu teknik pengambilan data pelatihan secara acak dengan pengembalian. Kemudian, ketika proses pembentukan cabang (*node splitting*) berlangsung, hanya sebagian kecil dari keseluruhan fitur yang dipilih secara acak untuk menentukan titik pemisahan yang paling optimal. Pendekatan ini dilakukan guna meminimalkan korelasi antar pohon serta meningkatkan kemampuan generalisasi model. Penentuan titik pemisahan terbaik dilakukan melalui perhitungan *information gain* yang berasal dari nilai entropi dataset, dengan rumus yang dapat dinyatakan sebagai berikut:

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

$Entropy(S)$ = Tingkat ketidakpastian pada himpunan data S

A = atribut atau fitur yang digunakan untuk pemisahan data

p_i = Probabilitas kemunculan kelas ke - i dalam dataset

S_i = Subset data hasil dari pemisahan berdasarkan atribut A

Dalam penelitian ini, data hasil ekstraksi fitur GLCM dibagi menjadi dua bagian, yaitu 80% data latih dan 20% data uji. Pelatihan dilakukan dengan membangun beberapa pohon keputusan untuk mendukung proses analisis data menggunakan fitur *contrast*, *energy*, *homogeneity*, dan *correlation* sebagai variabel prediktor. Hasil klasifikasi ditentukan melalui proses *majority voting*, di mana kelas yang memperoleh suara terbanyak dari seluruh pohon akan menjadi hasil akhir prediksi^[8]. *Random Forest* memiliki keunggulan dalam mengatasi masalah *overfitting*, mampu menangani data dengan dimensi yang besar, serta memberikan penilaian terhadap tingkat kepentingan setiap fitur pada proses klasifikasi. Dengan karakter yang adaptif dan efisien, algoritma ini cocok digunakan untuk analisis citra medis, termasuk dalam mengklasifikasikan penyakit Flu Singapura berdasarkan tekstur citra yang diperoleh.

2.6. Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja algoritma *Random Forest* dalam mengklasifikasikan citra lesi Flu Singapura (HFMD). Penilaian ini menggunakan *Confusion Matrix*, yang menggambarkan jumlah prediksi benar dan salah dari hasil klasifikasi. Maka diperoleh empat komponen utama, yaitu :

1. Akurasi (Accuracy)

$$accuracy = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Keseluruhan\ data}$$

2. Presisi (Precision)

$$precision = \frac{TP}{TP + FP}$$

3. Recall

$$recall = \frac{TP}{TP + FN}$$

4. F1-Score

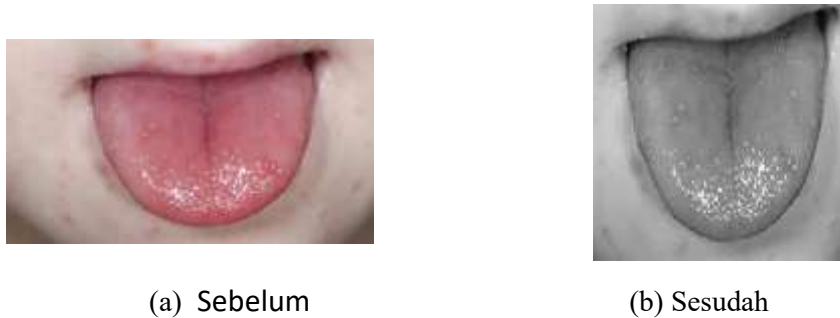
$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3. HASIL DAN PEMBAHASAN

3.1. Hasil *Preprocessing* Data

Dataset yang digunakan terdiri atas citra kulit yang menunjukkan indikasi penyakit Flu Singapura (HFMD), yang terbagi menjadi dua kategori yaitu positif dan negatif. Total data awal berjumlah 463 citra, terdiri atas 240 citra positif dan 223 citra negatif. Pre-processing dilakukan untuk meningkatkan kualitas citra sebelum tahap ekstraksi fitur dan klasifikasi. Tahapan ini meliputi penyeragaman ukuran citra menjadi 128 x 128 piksel, konversi ke skala abu-abu (*grayscale*), normalisasi nilai intensitas piksel, serta penghapusan citra yang buram, duplikat, atau tidak relevan.

Setelah proses *pre-processing* selesai, jumlah data yang layak digunakan berkurang menjadi 375 citra, terdiri dari 190 citra positif dan 185 citra negatif. Hasil ini menunjukkan bahwa tahapan *preprocessing* berhasil menyaring citra dengan kualitas terbaik untuk proses analisis selanjutnya. Contoh hasil *preprocessing* citra positif ditunjukkan pada Gambar 1.



Gambar 1. Contoh hasil *preprocessing* citra positif HFMD (a) sebelum, (b) sesudah

3.2. Hasil Ekstraksi Fitur Menggunakan *Gray Level Co-occurrence Matrix (GLCM)*

Tahap ekstraksi fitur dilakukan untuk memperoleh karakteristik tekstur dari citra hasil *preprocessing* menggunakan metode *Gray Level Co-occurrence Matrix (GLCM)*. Metode ini menganalisis hubungan spasial antar piksel dalam citra untuk menggambarkan pola tekstur yang mewakili permukaan kulit pada kasus Flu Singapura (HFMD). Empat parameter kunci yang digunakan dalam GLCM adalah *contrast*, *energy*, *homogeneity*, dan *correlation*, yang masing-masing menggambarkan tingkat perbedaan intensitas, keteraturan pola, keseragaman, serta hubungan linear antar piksel. Nilai-nilai ini menjadi representasi numerik dari citra yang kemudian digunakan sebagai masukan pada tahap klasifikasi. Tabel 1 memperlihatkan hasil perhitungan keempat parameter, yang menggambarkan nilai fitur citra positif dan negatif setelah melalui proses ekstraksi.

Filename	Contrast	Energy	Homogeneity	Correlation	Label
hfmd_rash_(108)	42.765	0.036	0.3131	0.973	positif
hfmd_rash_(107)	151.984	0.196	0.319	0.983	positif
hfmd_rash_(109)	108.914	0.021	0.228	0.975	positif
hfmd_rash_(109)	108.914	0.021	0.228	0.975	positif
hfmd_rash_(102)	267.217	0.018	0.187	0.968	positif
hfmd_rash_(106)	22.045	0.036	0.349	0.991	positif

hfmd_rash_(11)	75.494	0.025	0.223	0.985	positif
hfmd_rash_(103)	175.638	0.120	0.421	0.987	positif
hfmd_rash_(10)	136.932	0.024	0.283	0.964	positif
hfmd_rash_(104)	191.409	0.018	0.169	0.933	positif
hfmd_rash_(1)	38.570	0.038	0.337	0.991	positif
hfmd_rash_(110)	261.226	0.020	0.208	0.968	positif
hfmd_rash_(101)	37.494	0.026	0.286	0.988	positif
hfmd_rash_(100)	78.302	0.023	0.217	0.976	positif
hfmd_rash_(105)	332.671	0.035	0.192	0.954	positif
Not-hfmd_rash_(7)	28.465	0.022	0.239	0.990	negatif
Not-hfmd_rash_(70)	28.103	0.036	0.243	0.934	negatif
Not-hfmd_rash_(788)	45.529	0.031	0.286	0.951	negatif
Not-hfmd_rash_(779)	20.200	0.038	0.359	0.984	negatif
Not-hfmd_rash_(798)	12.920	0.043	0.285	0.985	negatif
Not-hfmd_rash_(737)	169.417	0.033	0.231	0.828	negatif
Not-hfmd_rash_(738)	155.008	0.017	0.157	0.964	negatif
Not-hfmd_rash_(804)	138.888	0.034	0.296	0.953	negatif
Not-hfmd_rash_(789)	58.850	0.027	0.242	0.966	negatif
Not-hfmd_rash_(781)	75.183	0.019	0.202	0.982	negatif
Not-hfmd_rash_(803)	72.724	0.055	0.260	0.986	negatif
Not-hfmd_rash_(735)	417.874	0.015	0.139	0.934	negatif
Not-hfmd_rash_(77)	147.50	0.025	0.241	0.980	negatif
Not-hfmd_rash_(801)	52.737	0.030	0.259	0.963	negatif
...	...	...	...	...	...

Tabel 1. Hasil Ekstraksi Fitur dengan GLCM

3.2. Hasil Klasifikasi *Random Forest*

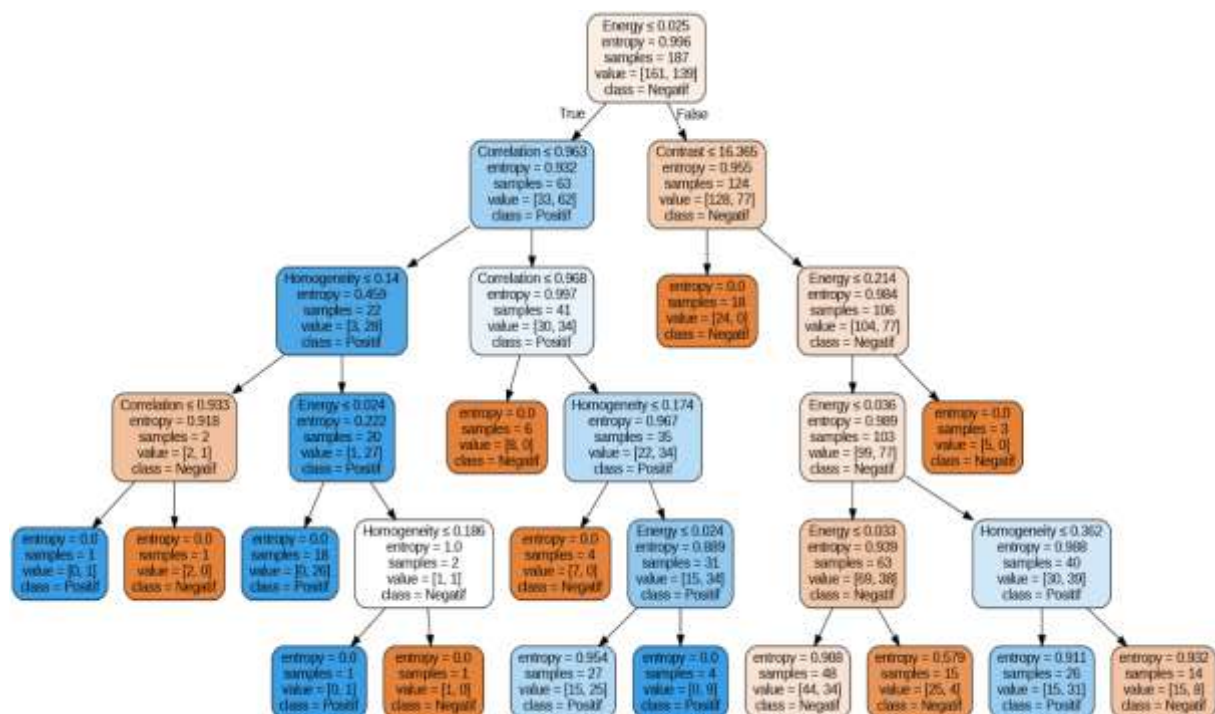
Tahap ini melibatkan proses klasifikasi dengan Algoritma Random Forest guna mengidentifikasi citra yang terindikasi Flu Singapura (HFMD) dari citra normal. Setelah melalui tahapan preprocessing, jumlah data yang digunakan sebanyak 375 citra, terdiri dari 190 citra positif dan 185 citra negatif. Dataset dibagi menjadi dua bagian, yaitu 80% digunakan sebagai data latih untuk pembelajaran model, dan 20% sisanya sebagai data uji untuk evaluasi kinerjanya. Pembagian ini dilakukan secara acak agar kedua kelompok data memiliki representasi data yang seimbang. Gambar 3 memperlihatkan distribusi data yang digunakan untuk pelatihan dan pengujian.



Gambar 3. Distribusi Data Pelatihan dan Pengujian

Dalam proses pelatihan, *Random Forest* membentuk 100 pohon keputusan (*decision trees*) dengan kedalaman maksimum lima tingkat. Pembentukan setiap pohon dilakukan melalui mekanisme bootstrap sampling dan *feature selection*, yang memungkinkan model menghasilkan variasi keputusan yang lebih stabil dan mengurangi risiko *overfitting*.

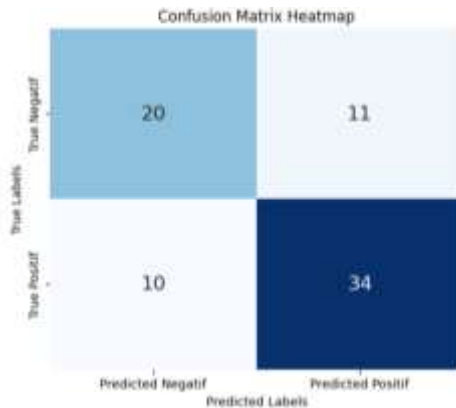
Untuk visualisasi, Gambar 4 menampilkan salah satu contoh pohon keputusan yang dihasilkan oleh model. Pohon tersebut menggambarkan proses pemisahan kelas berdasarkan nilai fitur tekstur hasil ekstraksi GLCM seperti *contrast*, *energy*, *homogeneity*, dan *correlation*. Pohon-pohon lain yang dihasilkan dalam ensemble bekerja bersama untuk menentukan hasil klasifikasi akhir dengan prinsip *majority voting*.



Gambar 4. Contoh pohon keputusan hasil pelatihan model *Random Forest*.

3.4. Hasil Evaluasi Model

Evaluasi dilakukan untuk menilai kinerja model *Random Forest* dalam mengklasifikasikan citra lesi kulit Flu Singapura (HFMD) menggunakan data uji. Hasil evaluasi ditunjukkan melalui *confusion matrix* pada Gambar 5.



Gambar 5. Hasil *Confusion Matrix*

Hasil evaluasi kinerja model, berupa nilai akurasi, *precision*, *recall*, dan *F1-score*) ditampilkan pada Gambar 6 berdasarkan *confusion matrix* sebelumnya

```

 **Hasil Evaluasi Model**
Akurasi: 0.7200
Precision: 0.7556
Recall: 0.7727
F1 Score: 0.7640

```

Gambar 6. Hasil Evaluasi Model

4. KESIMPULAN

Berdasarkan hasil penelitian, algoritma *Random Forest* terbukti mampu mengklasifikasikan citra lesi Flu Singapura (HFMD) dengan akurasi 72%, *precision* 75.56%, *recall* 77.27%, dan *F1-score* 76.40%. Hasil ini menandakan bahwa metode ini efektif untuk mendukung deteksi dini HFMD berbasis citra kulit. Ke depannya, peningkatan performa dapat dilakukan melalui penambahan variasi dataset, augmentasi citra, dan pengembangan sistem berbasis web atau mobile untuk penerapan praktis.

REFERENSI

- [1] L. M. Barus, H. G. Simanullang, and A. P. Silalahi, "Classification of English-Language Racial Hate Speech Using Random Forest", *J. Intell. Comput. Inf. Syst.*, vol. 1, no. 1, pp. 1–17, 2025.
- [2] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.
- [3] E. David, L. Tumanggor, H. G. Simanullang, and A. P. Silalahi, "Detection of AI-Generated Videos Using Local Binary Pattern and Support Vector Machine," vol. 18, no. 1, 2026.
- [4] S. Murni, D. Widiyanto, and catur nugrahaeni puspita Dewi, "Klasifikasi Citra Penyakit Daun Kopi Menggunakan Support Vector Machine (SVM) Dengan Seleksi Fitur Information Gain," *Senamika*, vol. 4, no. 2, pp. 700–709, 2022.
- [5] Mungki Astiningrum, Putra Prima Arhandi, and Nabilla Aqmarina Ariditya, "Identifikasi Penyakit Pada Daun Tomat Berdasarkan Fitur Warna Dan Tekstur," *J. Inform. Polinema*, vol. 6, no. 2, pp. 47–50, 2020, doi: 10.33795/jip.v6i2.320.
- [6] T. Tambunan, M. Yohanna, and A. P. Silalahi, "PENERAPAN METODE RANDOM FOREST DALAM MENDETEKSI BERITA HOAX," vol. 7, no. 2, pp. 301–306, 2023.