

Penerapan Algoritma Random Forest untuk Klasifikasi Jenis Daun Herbal

Sudiadi^{a1}, Meiriyama^{a2}

^aManajemen Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Indonesia

e-mail: 1Meiriyama@mdp.ac.id, 2Sudiadi@mdp.ac.id

Abstrak

Indonesia memiliki potensi yang besar dalam penyediaan sumberdaya tumbuhan obat atau tanaman herbal yang dapat dimanfaatkan dengan baik sebagai media pengobatan tradisional. Obat tradisional merupakan warisan lama turun temurun dari zaman dahulu, baik itu dalam bentuk ramuan maupun jamu. Tanaman obat merupakan spesies tanaman yang dipercaya sebagai obat alami tanpa kandungan kimia. Akan tetapi masih minimnya pengetahuan masyarakat mengenai jenis – jenis tanaman herbal. Oleh karena itu penelitian ini bertujuan untuk Klasifikasi Jenis Daun Herbal menggunakan Random Forest berdasarkan fitur HOG. Citra yang telah dipisah antara data latih dan data uji di ubah menjadi greyscale dan di resize menjadi 816x612 piksel, kemudian citra di ekstraksi menggunakan fitur HOG sehingga menghasilkan vektor sepanjang 1x3168. Algoritma Random Forest yang digunakan untuk klasifikasi daun herbal memiliki akurasi keseluruhan sebesar 85,33%.

Kata kunci: Daun, Herbal, HOG, Random Forest

Abstract

Indonesia has great potential in providing medicinal plant resources or herbal plants that can be used properly as a traditional medicine. Traditional medicine is an old heritage handed down from ancient times. Medicinal plants are plant species that are believed to be natural medicines without chemical ingredients. However, there is still a lack of public knowledge about the types of herbal plants. Therefore, this study aims to Classify Herbal Leaf Types using Random Forest based on the HOG feature. The image that has been separated between the training data and testing data is converted to greyscale and resized to 816x612 pixels, then the image is extracted using the HOG feature and produce 1x3168 vector. The Random Forest algorithm used for herbal leaf classification has an overall accuracy of 85.33%.

Keywords : Herb, HOG, Leaf, Random Forest

1. Pendahuluan

Indonesia memiliki potensi yang besar dalam penyediaan sumber daya tumbuhan obat atau tanaman herbal yang dapat dimanfaatkan dengan baik sebagai media pengobatan tradisional. Obat tradisional merupakan warisan lama turun temurun dari zaman dahulu, baik itu dalam bentuk ramuan maupun jamu. Tanaman obat merupakan spesies tanaman yang dipercaya sebagai obat alami tanpa kandungan kimia. Salah satu bagian tanaman obat (herbal) yang dapat digunakan sebagai bahan baku pembuatan obat tradisional adalah daun. Daun adalah bagian tanaman atau juga pohon yang menempel pada suatu batang atau cabang. Untuk mengetahui dan membedakan jenis daun tanaman obat maka dapat dibedakan melalui ciri daun tersebut berdasarkan bentuk, tekstur ataupun warna.

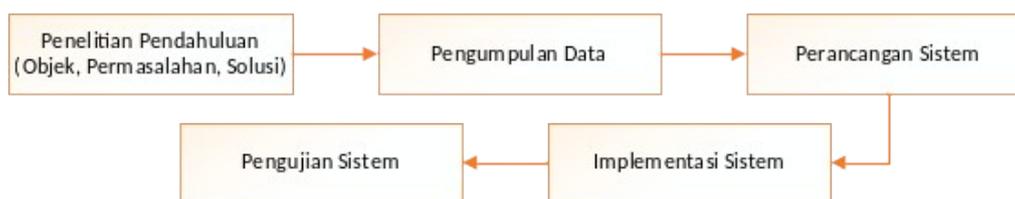
Sejumlah metode telah diusulkan untuk mengklasifikasikan tanaman berdasarkan citra daun. Penelitian yang dilakukan oleh [1] mengenai klasifikasi daun herbal berdasarkan fitur *invariant moment* dan fitur geometri yang kemudian diklasifikasi menggunakan algoritma KNN dan *Naïve Bayess*. Dalam penelitian [2] melakukan mengidentifikasi daun yang terdiri dari 1900 citra daun. Tiga langkah utama dalam penelitian tersebut antara lain praposes citra daun, ekstraksi fitur dengan menggunakan *Histogram of Oriented Gradients* (HOG) dan klasifikasi menggunakan KNN dan Jaringan Syaraf Tiruan. Hasil pengujian menunjukkan tingkat akurasi klasifikasi yaitu sebesar 97%. Ibrahim dkk [3] melakukan penelitian mengenai klasifikasi 10 jenis

daun tanaman herbal dengan menggunakan metode HOG, *Local Binary Pattern* (LBP), dan *Speeded Up Robust Features* (SURF) dengan algoritma klasifikasi yang digunakan adalah *Support Vector Machine* (SVM). Ratnawati dkk [4] dalam penelitiannya mengukur tingkat keparahan penyakit pada citra daun apel. Data yang digunakan sejumlah 467 citra daun apel yang dilakukan dalam beberapa tahapan proses yaitu pra-pengolahan citra, segmentasi citra menggunakan K-means clustering, ekstraksi fitur klasifikasi menggunakan metode *Random Forest*. Akurasi yang didapat adalah 75,3191%. Dalam Penelitian [5] mengklasifikasi citra bunga dengan menggunakan *Random forest* dengan fitur yang digunakan adalah *eccentricity*, *perimeter*, *metric* dan *area*. Berdasarkan hasil pengujian didapatkan bahwa metode *Random Forest* memiliki akurasi yang terbaik dibandingkan dengan SMO. Penelitian [6] yang melakukan klasifikasi penyakit Daun Padi dengan menggunakan algoritma *Random Forest* dan didapatkan akurasi sebesar 99,65%.

Berdasarkan penelitian terdahulu maka pada penelitian ini akan dilakukan klasifikasi tanaman herbal berdasarkan fitur bentuk dengan menggunakan metode HOG. Metode klasifikasi yang digunakan adalah *Random Forest*. Sebelum dilakukan tahapan pengelompokkan, maka terlebih dahulu dilakukan tahapan praproses citra dengan tujuan untuk mendapatkan citra masukan yang baik.

2. Metodologi

Berikut adalah beberapa tahapan yang dilakukan untuk mengklasifikasi jenis daun herbal dengan menggunakan HOG dan *Random Forest* yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Penelitian ini menggunakan dataset yang berjumlah 450 citra daun herbal yang terdiri dari 15 jenis daun yaitu daun sirih merah, daun bayam merah, daun mangkok, daun jambu biji, daun tapak dara, daun jarak, daun binahong, daun salam, daun kumis kucing, daun kersen, daun katuk, daun sirsak, daun kembang sepatu, daun urang aring, dan daun insulin. Dataset kemudian akan dibagi menjadi dua yaitu data latih dan data uji. Setiap jenis daun mempunyai ukuran 816 x 612 piksel. Berikut adalah jumlah data yang terdapat pada setiap jenis daun yang dapat dilihat pada Tabel 1.

Tabel 1. Dataset Daun Herbal

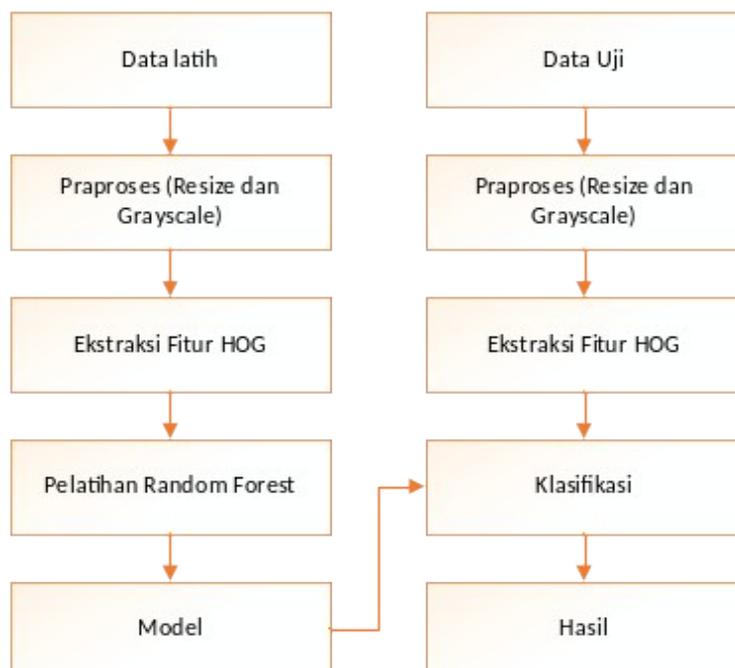
No	Jenis Daun	Data Latih	Data Uji
1	Daun Bayam Merah	20 Citra Daun	10 Citra Daun
2	Daun Binahong Merah	20 Citra Daun	10 Citra Daun
3	Daun Insulin	20 Citra Daun	10 Citra Daun
4	Daun Jambu biji	20 Citra Daun	10 Citra Daun
5	Daun Jarak	20 Citra Daun	10 Citra Daun
6	Daun Katuk	20 Citra Daun	10 Citra Daun
7	Daun Kembang Sepatu	20 Citra Daun	10 Citra Daun
8	Daun Kumis Kucing	20 Citra Daun	10 Citra Daun
9	Daun Mangkok	20 Citra Daun	10 Citra Daun
10	Daun Salam	20 Citra Daun	10 Citra Daun
11	Daun Seri	20 Citra Daun	10 Citra Daun
12	Daun Sirih Merah	20 Citra Daun	10 Citra Daun
13	Daun Sirsak	20 Citra Daun	10 Citra Daun
14	Daun Tapak Dara	20 Citra Daun	10 Citra Daun
15	Daun Urang Aring	20 Citra Daun	10 Citra Daun



Gambar 2. Jenis Daun Herbal

2.2 Perancangan Sistem

Pada tahapan ini dilakukan perancangan sistem untuk klasifikasi jenis daun herbal yang terdiri dari dua proses utama yaitu proses pelatihan dan proses pengujian yang dapat dilihat pada Gambar 3. Proses pelatihan menggunakan metode *Random Forest* yang menghasilkan sebuah model klasifikasi.



Gambar 3. Tahapan Perancangan Sistem

2.3 Implementasi Sistem

Setelah melakukan proses perancangan, pada tahap ini dilakukan implementasi terhadap rancangan yang telah dibuat kedalam program matlab.

2.4 Pengujian Sistem

Setelah melakukan proses implementasi dilanjutkan dengan proses pengujian terhadap data uji yang berjumlah 150 citra daun herbal. Pada tahapan ini didapatkan hasil berupa *Accuracy*, *Precision*, dan *Recall*.

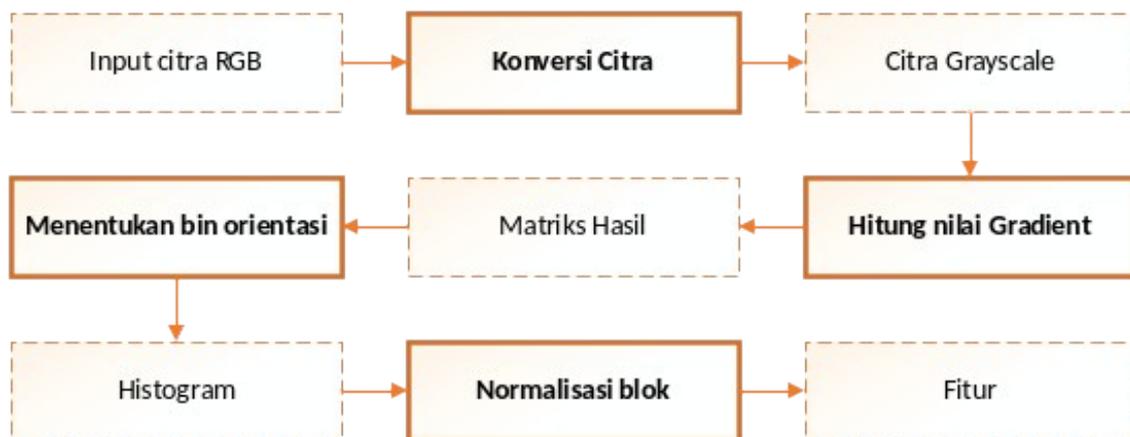
3. Studi Literatur

3.1 Daun Tanaman Herbal

Tanaman herbal banyak digunakan sebagai bahan baku obat tradisional dan jamu. Berdasarkan Riset Kesehatan Dasar (Riskesdas) 2010 bahwa persentasi penduduk Indonesia yang pernah mengonsumsi jamu herbal adalah sebanyak 59,12% yang terdapat pada kelompok umur di atas 15 tahun dan 95,60% merasakan manfaatnya penggunaan tanaman herbal sebagai pengobatan alternatif atau tradisional. Banyak khasiat dan manfaat yang dapat diperoleh dari penggunaan daun tanaman herbal yaitu salah satunya dapat meningkatkan imun tubuh sehingga dapat menjadi alternatif untuk menjaga Kesehatan [7].

3.2 Histogram of Oriented Gradient

Terdapat 4 langkah utama pada metode Histogram of Oriented Gradient yang dapat dilihat pada Gambar 4.

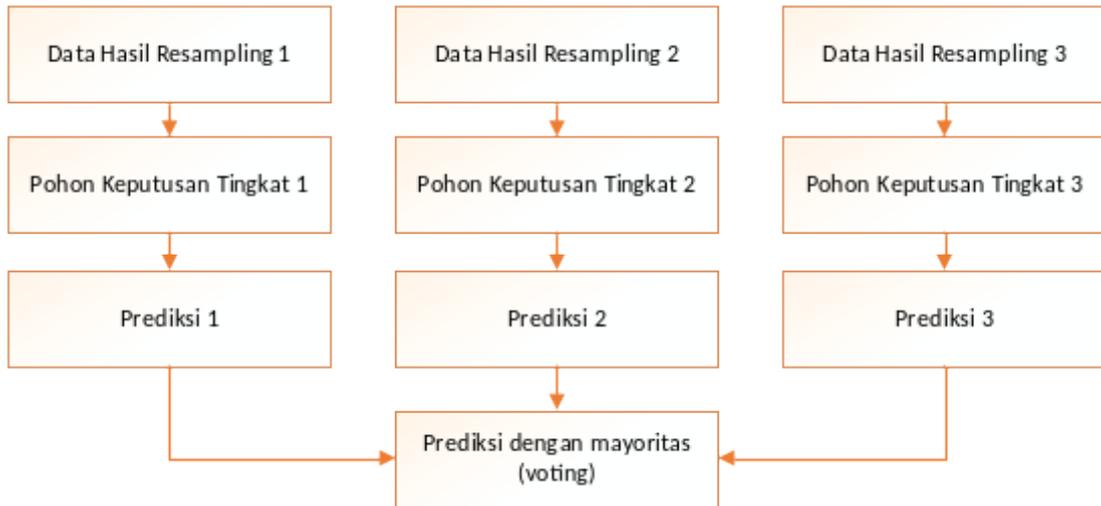


Gambar 4. Tahapan HOG [8]

- Konversi Citra yaitu mengubah citra RGB (Red, Green, Blue) menjadi citra grayscale.
- Perhitungan nilai gradient yang dilakukan untuk setiap piksel pada citra grayscale dengan membagi citra menjadi bentuk cell
- Menentukan nilai bin orientasi
- Melakukan normalisasi nilai setiap bin orientasi yang kemudian akan menghasilkan fitur

3.3 Random Forest

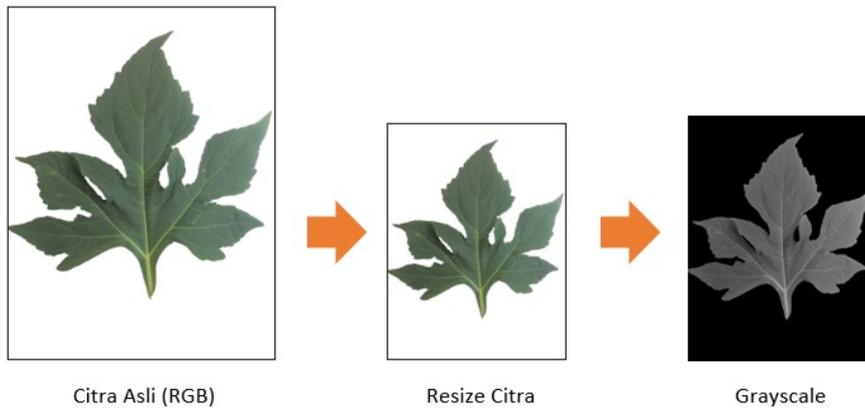
Random forest adalah algoritma yang menggunakan pendekatan decision trees dari keseluruhan decision trees ini akan membentuk hutan atau 'forest'. Keunggulan dari random forest diantaranya yaitu memiliki fitur seleksi yang dapat mengambil dan memilih fitur terbaik secara acak untuk meningkatkan performa pada klasifikasi dan meningkatkan akurasi walau terdapat *noise* [4]. Gambar 5 menunjukkan algoritma random forest.



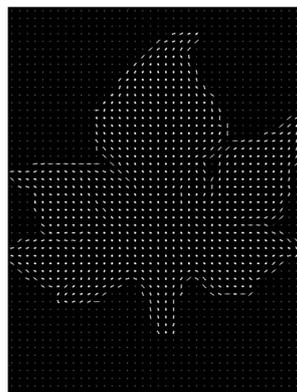
Gambar 5. Algoritma Random Forest [4]

4. Hasil dan Pembahasan

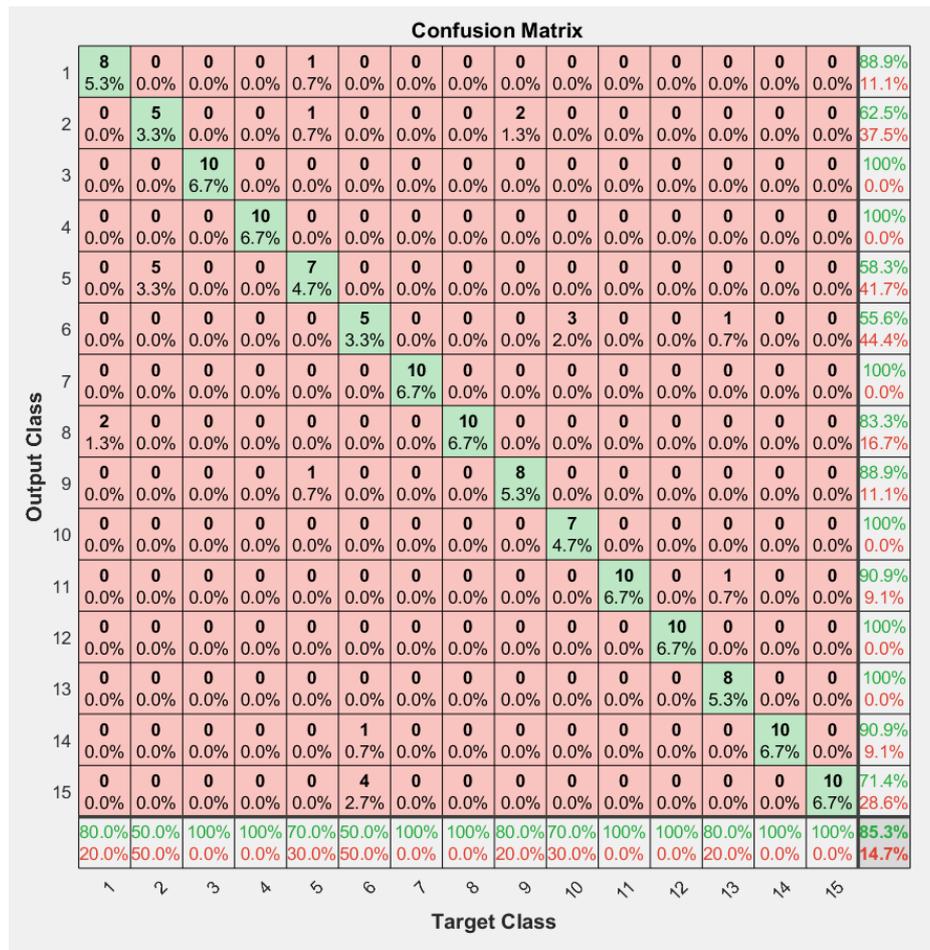
Praproses citra daun adalah langkah pertama di mana ukuran citra diubah (resize) menjadi 816 x 612 pixel kemudian dilakukan grayscale, seperti yang ditunjukkan pada Gambar 6. Kemudian citra grayscale akan diekstraksi dengan menggunakan HOG. Hasil ekstraksi HOG untuk satu citra daun tanaman herbal adalah vektor yang berukuran 1x3168. Kemudian dilakukan klasifikasi dengan menggunakan *random forest*.



Gambar 6. Praproses Citra Daun Herbal



Gambar 7. Ekstraksi HOG



Gambar 8. Confusion Matrix

Gambar 8 adalah Confusion matriks yang didapatkan dari hasil pengujian terhadap 150 citra daun herbal. Dari hasil Perhitungan Confusion Matrix yang dilakukan untuk klasifikasi 15 jenis Daun Herbal dihasilkan nilai yang ditunjukkan oleh Tabel 2 sebagai berikut :

Tabel 1. Dataset Daun Herbal

No	Jenis Daun	Accuracy (%)	Precision	Recall
1	Daun Bayam Merah	98	0,89	0,8
2	Daun Binahong Merah	94,67	0,69	0,5
3	Daun Insulin	100	1,0	1,0
4	Daun Jambu biji	100	1,0	1,0
5	Daun Jarak	94,67	0,58	0,7
6	Daun Katuk	94	0,56	0,5
7	Daun Kembang Sepatu	100	1,0	1,0
8	Daun Kumis Kucing	98,67	0,83	1,0
9	Daun Mangkok	98	0,89	0,8
10	Daun Salam	98	1,0	0,7
11	Daun Seri	99,33	0,91	1,0
12	Daun Sirih Merah	100	1,0	1,0
13	Daun Sirsak	98,67	1,0	0,8
14	Daun Tapak Dara	99,33	0,91	1,0
15	Daun Urang Aring	97,33	0,71	1,0

Berdasarkan hasil perhitungan menunjukkan bahwa model klasifikasi daun herbal dengan menggunakan *random forest* berhasil melakukan klasifikasi secara baik terhadap kelas Daun Insulin, Daun Jambu Biji dan Daun Sirih Merah yang dibuktikan dengan tingkat akurasi 100% Nilai precision terkecil adalah 0,56 yaitu berada pada kelas Daun Katuk yang menandakan bahwa model klasifikasi lebih banyak mengklasifikasikan untuk data false positive. Nilai recall terkecil adalah 0,5 yaitu berada pada kelas Daun katuk dan Daun Binahong Merah yang menandakan bahwa model klasifikasi lebih banyak mengklasifikasikan untuk data false negative. Dan secara keseluruhan akurasi yang dihasilkan adalah 85,33%.

5. Kesimpulan

Berdasarkan hasil pengujian dan pembahasan dapat diambil kesimpulan sebagai berikut:

- a. Ekstraksi fitur HOG pada daun herbal menghasilkan fitur yang berukuran 1x3168 untuk setiap citra daun herbal
- b. Dari hasil pengujian tersebut didapatkan akurasi secara keseluruhan yaitu 85,3%, rata – rata recall sebesar 0,85 dan rata – rata presisi sebesar 0,86 sehingga dapat dikatakan bahwa sistem telah mampu melakukan klasifikasi dengan cukup baik.

References

- [1] F. Liantoni and H. Nugroho, "Klasifikasi Daun Herbal Menggunakan Metode Naïve Bayes Classifier Dan Knearest Neighbor," *J. Simantec*, vol. 5, no. 1, pp. 9–16, 2015.
 - [2] P. Sharma, A. Aggarwal, A. Gupta, and A. Garg, "Leaf identification using HOG, KNN, and neural networks," in *Lecture Notes in Networks and Systems*, vol. 56, Springer, 2019, pp. 83–91.
 - [3] Z. Ibrahim, N. Sabri, and N. N. A. Mangshor, "Leaf recognition using texture features for herbal plant identification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 1, pp. 152–156, Jan. 2018, doi: 10.11591/ijeecs.v9.i1.pp152-156.
 - [4] L. Ratnawati and D. R. Sulistyaningrum, "Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel," *J. Sains dan Seni ITS*, vol. 8, no. 2, 2020, doi: 10.12962/j23373520.v8i2.48517.
 - [5] P. Rosyani, S. Saprudin, and R. Amalia, "Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 132, 2021, doi: 10.26418/justin.v9i2.44120.
 - [6] *et al.*, "Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram," *J. Komputasi*, vol. 10, no. 1, 2022, doi: 10.23960/komputasi.v10i1.2961.
 - [7] K. S. Artini and W. Veranita, "Tanaman Herbal Untuk Meningkatkan Sistem Imun Tubuh : Literature Review," *J. Farmasetis*, vol. 10, no. 1, pp. 15–20, 2021.
 - [8] Y. Yohannes, Y. P. Sari, and I. Feristyani, "Klasifikasi Wajah Hewan Mamalia Tampak Depan Menggunakan k-Nearest Neighbor Dengan Ekstraksi Fitur HOG," *J. Tek. Inform. dan Sist. Inf.*, vol. 5, no. 1, 2019, doi: 10.28932/jutisi.v5i1.1584.
-