

Klasifikasi Penyakit Pada Tanaman Kopi Menggunakan K-Nearest Neighbor dengan dioptimasi dengan Genetic Algorithm

1st M Alif Naufal Yasin
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia

alifnaufalyasin@student.telkomuniversity.ac.id

2nd Wikky Fawwaz Al Maki
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia

wikkyfawwaz@telkomuniversity.ac.id

Abstrak

Daun kopi merupakan salah satu bagian yang paling mudah untuk dideteksi Ketika tanaman kopi terserang penyakit. Banyak sekali faktor yang dapat mengakibatkan tanaman kopi menjadi sakit, beberapa faktor penyebabnya yaitu hama, karat daun, jamur daun dan juga bercak daun. Dengan memanfaatkan image processing dapat membantu mengklasifikasikan penyakit melalui citra daunnya, sehingga dapat dengan tepat diberikan penanganan pada tanaman yang sakit tersebut. Dengan melakukan ekstraksi fitur pada fitur warna menggunakan Color Moment. Kemudian dilakukannya klasifikasi menggunakan metode non deep learning K-Nearest Neighbor dan dilakukan optimasi dengan menggunakan Genetic Algorithm, maka didapatkan nilai k yang paling optimal sehingga nilai akurasi yang didapatkan adalah nilai akurasi maksimal yang bisa dihasilkan oleh KNN. Hasil tersebut didapatkan akurasi terbaik dengan menggunakan color moment pada ruang warna RGB yaitu sebesar 78.9% dengan menggunakan klasifikasi KNN+GA.

I. PENDAHULUAN

A. Latar Belakang

Kopi merupakan salah satu komoditas yang dibudidayakan oleh banyak negara. Kopi menjadi komoditas terbesar kedua setelah minyak bumi [1]. Bagi sebagian negara, kopi menjadi salah satu pendapatan terbesar bagi negara tersebut. Kopi yang baik didapatkan dari tanaman kopi yang sehat. Tanaman yang tidak sehat akan menghasilkan kopi yang kurang berkualitas seperti rasa, aroma atau tingkat keasamannya berkurang, atau rasanya yang menjadi lebih pahit [2]. Banyak factor yang dapat mengakibatkan tanaman kopi menjadi kurang sehat. Hama merupakan salah satu penyebab dari tanaman kopi yang kurang sehat. Selain hama, ada beberapa penyebab lainnya yang dapat merusak tanaman kopi, yaitu karat daun (rust leaf), jamur daun (phoma) dan juga munculnya bercak daun (Cercospora).

Daun kopi merupakan salah satu bagian yang dapat mengidentifikasi apakah tanaman tersebut

Kata kunci : image processing, color moment, k-nearest neighbor, genetic algorithm, daun kopi

Abstract

Coffee leaves are one of the most accessible parts to detect when the coffee plant is attacked by disease. Many factors can cause coffee plants to become sick. Some of the contributing factors are pests, leaf rust, leaf fungus, and leaf spot. By utilizing image processing, it can help classify diseases through the image of the leaves to provide proper treatment for these plants by performing color feature extraction using Color Moment. Then the classification uses the non-deep learning K-Nearest Neighbor method, which is optimized using Genetic Algorithm on its k value. The most optimal k value is obtained so that the accuracy value obtained is also higher. The results obtained the best accuracy using the color moment in the RGB color space is 78.9% using the KNN+GA classification.

Keywords: image processing, color moment, k-nearest neighbor, genetic algorithm, coffee leaves

sehat atau sakit. Daun kopi akan berubah warna ataupun munculnya bercak pada daun. Oleh karena itu, penyakit pada tanaman kopi dapat diklasifikasi dengan menggunakan citra daun kopi. Sebagai contoh, pada tahun 2020 terdapat penelitian klasifikasi penyakit tanaman kopi dengan menggunakan metode Deep Learning yaitu Convolutional Neural Network (CNN) dengan arsitektur ResNet50. Dari penelitian tersebut didapatkan akurasi sebesar 97.07% [3]. Berdasarkan latar belakang diatas, penulis mengusulkan system klasifikasi penyakit pada daun tanaman kopi dengan menggunakan metode non deep learning yaitu K-Nearest Neighbor yang dioptimasi menggunakan Genetic Algorithm. Kemudian menggunakan metode ekstraksi ciri pada warna menggunakan Color Moment dengan kategori kelas Sehat (healthy), Karat Daun (Rust), Jamur Daun (Phoma) dan juga bercak daun (Cercospora).

B. Topik dan Batasannya

Pada penelitian ini, akan dilakukannya klasifikasi penyakit pada tanaman kopi dengan menggunakan citra daun kopi. Metode yang digunakan adalah Non Deep Learning dengan menggunakan KNN yang dioptimasi dengan GA. Dataset yang digunakan pada penelitian ini memiliki 3 jenis penyakit yaitu Karat Daun (Rust), Jamur Daun (Phoma) dan juga bercak daun (Cercospora), ditambah dengan 1 daun sehat, sehingga total menjadi 4 kelas. Data citra yang digunakan yaitu citra symptom dari daun tanaman kopi dengan resolusi citra yang berbeda-beda.

C. Tujuan

Penelitian ini bertujuan untuk membangun system klasifikasi penyakit pada daun tanaman kopi menggunakan metode ekstraksi color moment, serta metode klasifikasi K-Nearest Neighbor yang dioptimasi menggunakan Genetic Algorithm.

D. Organisasi Tulisan

Laporan hasil penelitian terdiri dari lima bagian. Bagian pertama yaitu pendahuluan. Bagian kedua berisi tentang penelitian yang berkaitan dengan penelitian ini. Bagian ketiga menjelaskan tentang sistem yang dibangun pada penelitian ini. Bagian keempat menjelaskan evaluasi dari hasil yang didapatkan oleh sistem yang telah dibangun. Pada bagian kelima berisi kesimpulan secara keseluruhan dari penelitian yang telah dilakukan.

II. KAJIAN TEORI

Berdasarkan penelitian yang dilakukan oleh José G.M. Esgario, Renato A. Krohling dan José A. Ventura yang berjudul “*Deep learning for classification and severity estimation of coffee leaf biotic stress*” dengan menggunakan metode deep learning yaitu dengan CNN berarsitektur ResNet50, didapatkan akurasi keseluruhan sebesar 97.07% [3].

Berdasarkan penelitian yang telah dilakukan pada tahun 2020 yang berjudul “*Implementation of Artificial Neural Networks Using NVIDIA Digits and OpenCV for Coffee Rust Detection*” dengan menggunakan metode Open CV Neural Network untuk proses trainingnya, maka didapatkan akurasi rata-rata sebesar 83%. Dataset yang digunakan merupakan dataset yang dikumpulkan dari kebun kopi dan divalidasi oleh presiden Asosiasi Produsen Kopi Honduras melalui sebuah wawancara pribadi. Dataset memiliki resolusi yang berbeda-beda sehingga perlu dilakukannya penyesuaian resolusi menjadi 256x256 piksel [4].

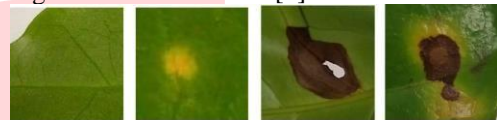
Berdasarkan penelitian yang dilakukan pada tahun 2020 yang berjudul “*Image processing based Automatic Plant Disease Detection and Stem Cutting Robot*” dengan menggunakan segmentasi untuk mendapatkan citra dari daun yang sakit, kemudian dilakukannya ekstraksi ciri pada warna dan diklasifikasikan menggunakan Neural Network.

Akurasi yang didapatkan pada penelitian ini yaitu sebesar 79%. Dataset yang digunakan yaitu citra daun kopi dengan total 600 citra yang dibagi menjadi 3 kelas penyakit dan 1 yang sehat dengan tiap kelas sejumlah 150 citra [5].

III. METODE

A. Dataset

Dataset yang digunakan pada penelitian ini yaitu citra symptom dari daun tanaman kopi. Dataset yang digunakan memiliki 4 kelas dengan 3 kelas penyakit dan 1 kelas sehat dengan total dataset berjumlah 1669 citra. Dataset didapatkan dari penelitian yang dilakukan oleh José G.M. Esgario, Renato A. Krohling dan José A. Ventura [3].



GAMBAR 1 Dataset Citra Symptom

B. Prapemrosesan

Prapemrosesan data dilakukan untuk memformat gambar, mengurangi komponen gambar yang tidak diinginkan, serta dapat meningkatkan kualitas gambar untuk dapat menghasilkan nilai akurasi yang lebih tinggi [6]. Pada penelitian ini, akan dilakukannya penskalaan ulang resolusi dari seluruh citra agar memiliki resolusi citra yang sama. Kemudian dilakukannya perubahan ruang warna RGB menjadi *grayscale* untuk pemrosesan pada ekstraksi ciri GLCM dan juga merubah ruang warna RGB menjadi HSV untuk pemrosesan ekstraksi ciri dengan metode color moment.



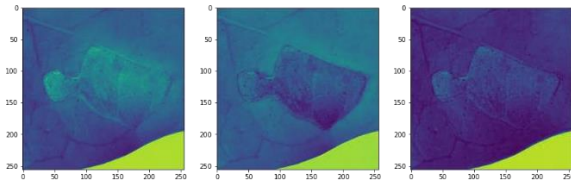
GAMBAR 2 Sebelum Dilakukan Skala Ulang

GAMBAR 3 Setelah Dilakukan Skala Ulang

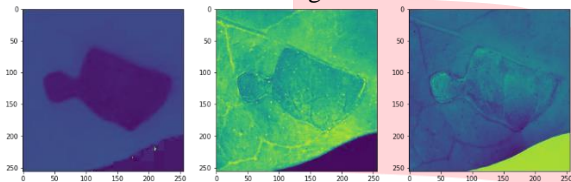
C. Metode Ekstraksi Ciri Color Moment

Color moment merupakan sebuah metrik yang dapat digunakan untuk membedakan gambar berdasarkan dari kesamaan warna antar gambar [7].

Color moment didasarkan pada konsep distribusi warna yang biasa disebut sebagai distribusi probabilitas. Kemudian, jika terdapat warna dalam suatu gambar mengikuti distribusi probabilitas tertentu, hal tersebut dapat digunakan untuk mengidentifikasi secara unik sebuah gambar [8]. Pada penelitian ini menggunakan dua ruang warna untuk dilakukannya ekstraksi ciri dengan color moment, yaitu RGB dan HSV.



GAMBAR 4 Ruang Warna RGB



GAMBAR 5 Ruang Warna HSV

Untuk setiap ruang warna, kami menggunakan tiga jenis distribusi warna yang sering digunakan untuk ekstraksi ciri pada warna yaitu: *mean*, *standard deviation*, dan *skewness* [7], [8].

a. Mean

Nilai dari Mean merupakan nilai rata-rata piksel pada setiap channel warna. Berikut adalah rumus untuk mendapatkan nilai mean.

$$E_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad (1)$$

Dimana E_i , N , dan P_{ij} adalah nilai mean pada ruang warna i , total piksel, dan piksel j pada ruang warna i .

b. Standard Deviation

Standard deviation adalah suatu ukuran dari persebaran angka atau warna pada gambar. Berikut adalah rumus untuk mendapatkan nilai standard deviation.

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^2} \quad (2)$$

Dimana E_i , N , P_{ij} , dan σ_i adalah nilai mean pada ruang warna i , total piksel, piksel j pada ruang warna i , dan nilai standar deviasi pada ruang warna i .

c. Skewness

Skewness merupakan representasi dari ketidak seimbangan dan asimetri dari rata-rata distribusi data/warna. Berikut

adalah rumus untuk mendapatkan nilai skewness.

$$S_i = \frac{\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^3}{\sigma_i^3} \quad (3)$$

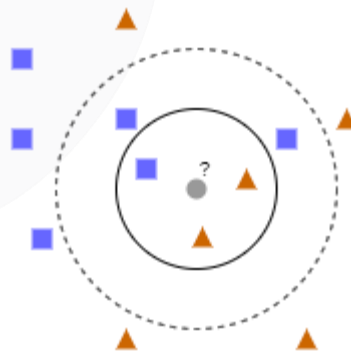
Dimana E_i , N , P_{ij} , σ_i^3 , dan S_i adalah nilai mean pada ruang warna i , total piksel, piksel j pada ruang warna i , nilai standar deviasi pada ruang warna i , dan nilai skewness pada ruang warna i .

D. Metode Seleksi Fitur Backward Elimination

Backward elimination merupakan bagian dari metode *wrapper* yang merupakan kebalikan dari *forward selection*. Pada proses ini, semua variabel input dipilih terlebih dahulu, kemudian variabel yang paling tidak penting akan dihilangkan satu persatu hingga tersisa variabel input yang penting [9]. Variabel hasil dari seleksi fitur akan menghasilkan nilai akurasi yang lebih baik karena data yang digunakan hanya variabel berpengaruh.

E. Metode Klasifikasi

Pada penelitian ini, kami menggunakan K-Nearest Neighbor untuk klasifikasinya yang kemudian dioptimasi dengan menggunakan Genetic Algorithm. KNN merupakan tipe supervised learning yang biasa digunakan untuk memecahkan masalah tentang pengenalan pola [10]. KNN juga merupakan metode klasifikasi berbasis jarak antar tetangga. Data akan diklasifikasikan berdasarkan mayoritas dari tetangganya. Gambar 6 merupakan ilustrasi bagaimana KNN membuat keputusan dengan berdasarkan ketetanggaannya, keputusan yang diambil akan berbeda antara tiga tetangga dengan lima tetangga.

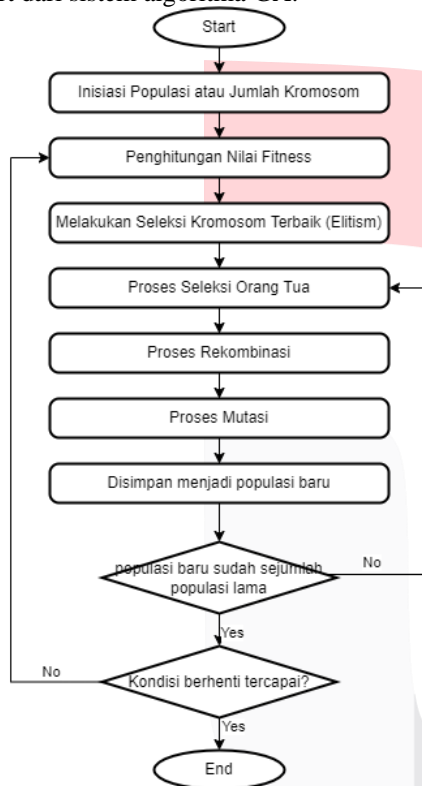


Gambar 6 Proses Keputusan Algoritma KNN

Genetic Algorithm terinspirasi dari teori evolusi Darwin, dimana keberlangsungan makhluk hidup yang lebih kuat serta gen mereka disimulasikan [11]. GA merupakan algoritma berbasis populasi. Setiap individu akan merepresentasikan solusi yang didapatkan melalui kromosom gen dan setiap parameternya. GA akan melakukan evaluasi terhadap nilai fitness dari setiap individu pada populasi.

$$k = \sum_{i=1}^{length} 2^{i-1} \times X_i \tag{10}$$

Persamaan (10) merupakan formula untuk mendapatkan nilai k untuk setiap individu pada algoritma KNN-GA, dimana nilai X_i merupakan nilai kromosom ke-i pada sebuah individu. Nilai k pada KNN akan dioptimasi dengan menggunakan GA, dan hasil dari klasifikasi tersebut akan menjadi nilai fitness dari setiap individu pada GA. Sehingga nilai fitness terbaik akan menghasilkan nilai k dengan hasil klasifikasi KNN terbaik. Gambar 7 merupakan flowchart dari sistem algoritma GA.



Gambar 7 Flowchart algoritma GA

F. Metode Evaluasi

Hasil dari klasifikasi akan dilakukannya evaluasi. Nilai akurasi pada penelitian ini diperlukan untuk menghitung performa dari model atau sistem yang sudah dibuat. Akurasi merupakan rasio dari jumlah prediksi yang benar dari jumlah seluruh data, hal tersebut dapat dihitung dengan persamaan (11). Selain akurasi, terdapat juga nilai precision, recall serta F1-score yang dapat dihitung dengan persamaan (12 - 14).

$$Akurasi = \frac{(TP+TN)}{(TP+TN + FP + FN)} \tag{11}$$

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

(12)

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

(13)

$$score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{14}$$

F1 –

IV. HASIL DAN PEMBAHASAN

A. Skenario Pengujian

Penelitian ini menggunakan ekstraksi fitur pada warna menggunakan color moment. Kemudian dilakukannya klasifikasi dengan menggunakan KNN yang kemudian dioptimasi dengan menggunakan PSO. Dataset yang digunakan akan dibagi menjadi 2 bagian, 75% data latih dan 25% data uji. Pada KNN formula perhitungan jarak antar tetangga yang digunakan yaitu formula minkowski dengan nilai k yaitu 5. Pada seleksi fitur dengan backward elimination, menggunakan parameter P-value yaitu 5%. Jumlah individu pada sebuah populasi yaitu sebanyak 50 individu yang Panjang kromosom dari tiap individu yaitu 11 digit dengan representasi biner. Untuk peluang terjadinya proses mutasi dan rekombinasi yaitu 25% dan 85%. Kemudian untuk kondisi pencarian individu terbaik akan berhenti jika telah tercapai nilai fitness atau akurasi lebih besar dari 90% atau jika sudah mencapai generasi ke 50.

B. Hasil Pengujian

Dilakukannya ekstraksi fitur dengan menggunakan color moment yang kemudian dilakukan seleksi fitur dengan menggunakan backward elimination. Pada Tabel 1, terlihat bahwa terdapat beberapa nilai P-value yang melebihi dari yang sudah ditentukan sebelumnya yaitu 5%. Pada RGB terdapat pada fitur X9 atau nilai skewness pada channel warna biru yang memiliki nilai P-value 9.3%. Kemudian fitur X9 tersebut dikeluarkan dan dilakukan pengecekan Kembali namun tidak ada perbedaan dari hasil akurasi yang didapatkan sehingga nilai X9 dikembalikan Kembali dan semua fitur digunakan pada ruang warna RGB. Lain halnya dengan ruang warna HSV. Pada fitur X3 atau nilai Mean pada channel warna V atau value memiliki nilai P-value sebesar 87.5% sehingga fitur X3 akan

dikeluarkan dan dilakukannya pengecekan Kembali. Hasil akurasi yang didapatkan meningkat saat fitur X3 dikeluarkan, sehingga fitur yang digunakan pada ruang warna HSV adalah seluruh fitur dengan fitur X3 dikeluarkan.

TABEL 1 Hasil Backward Elimination

Fitur	P-value	
	RGB	HSV
X1 (R/H Mean)	0.0	0.0
X2 (G/S Mean)	0.0	0.025
X3 (B/V Mean)	0.001	0.875
X4 (R/H Standard Deviation)	0.0	0.0
X5 (G/S Standard Deviation)	0.0	0.0
X6 (B/V Standard Deviation)	0.04	0.0
X7 (R/H Skewness)	0.002	0.015
X8 (G/S Skewness)	0.0	0.0
X9 (B/V Skewness)	0.093	0.01

Kemudian Dilakukannya percobaan untuk fitur warna yaitu menggunakan color moment dengan ruang warna RGB dan HSV. Hasil yang didapatkan pada TABEL yaitu pada ruang warna RGB, didapatkan angka akurasi sebesar 78.5% saat menggunakan KNN dan jika dilakukan optimasi dengan GA maka didapatkan akurasi sebesar 78.9%. Begitu pula pada ruang warna HSV dengan akurasi sebesar 75.1% dengan menggunakan KNN dan akurasi sebesar 75.8% dengan KNN yang dioptimasi menggunakan GA.

TABEL 2 Hasil Color Moment

Ruang Warna	Metode Klasifikasi	Akurasi	Precisi	Recall	F1-Score
RGB	KNN	0.785	0.77	0.76	0.77
	KNN+GA	0.789	0.77	0.76	0.76
HSV	KNN	0.751	0.74	0.72	0.72
	KNN+GA	0.758	0.74	0.72	0.73

Kemudian dilakukannya pengecekan nilai k pada RGB dan juga HSV dengan nilai k yaitu 1 hingga 15 untuk memastikan bahwa hasil yang didapatkan oleh KNN+GA merupakan hasil akurasi terbaik dari KNN. Pada Tabel 3, hasil akurasi yang didapatkan

oleh KNN+GA untuk ruang warna RGB yaitu dengan k = 9 dan hasil tersebut merupakan nilai akurasi terbesar. Kemudian pada Tabel 4, hasil akurasi yang didapatkan oleh KNN+GA pada ruang warna HSV yaitu k = 6 atau 10 dan hasil tersebut juga merupakan nilai akurasi terbesar. Nilai k pada ruang warna RGB dan juga HSV berbeda, oleh karena itu GA membantu untuk mengoptimasi dalam menemukan nilai k terbaik pada fitur yang digunakan.

TABEL 3 Pengecekan Nilai k Pada Color Moment RGB

Nilai k	Akurasi	Precisi	Recall	F1-Score
1	0.761	0.75	0.75	0.74
2	0.763	0.76	0.74	0.74
3	0.775	0.76	0.76	0.76
4	0.78	0.77	0.75	0.76
5	0.785	0.77	0.76	0.77
6	0.787	0.77	0.76	0.76
7	0.785	0.77	0.76	0.76
8	0.78	0.77	0.74	0.75
9	0.789	0.77	0.75	0.76
10	0.78	0.76	0.74	0.75
11	0.77	0.76	0.73	0.74
12	0.768	0.75	0.73	0.73
13	0.766	0.75	0.72	0.73
14	0.766	0.75	0.72	0.73
15	0.761	0.75	0.72	0.73

TABEL 4 Pengecekan Nilai k Pada Color Moment HSV

Nilai k	Akurasi	Precisi	Recall	F1-Score
1	0.706	0.69	0.69	0.69
2	0.689	0.67	0.67	0.66
3	0.706	0.68	0.68	0.68
4	0.739	0.71	0.71	0.71
5	0.751	0.74	0.72	0.72
6	0.758	0.74	0.72	0.73
7	0.754	0.73	0.72	0.72
8	0.754	0.73	0.71	0.72
9	0.754	0.74	0.72	0.73
10	0.758	0.74	0.73	0.73
11	0.744	0.73	0.71	0.72
12	0.746	0.73	0.71	0.72
13	0.739	0.73	0.71	0.71
14	0.734	0.72	0.7	0.71
15	0.734	0.72	0.7	0.71

C. Analisis Hasil Pengujian

Berdasarkan hasil pada Tabel 2, Pada ruang warna RGB ataupun HSV didapatkannya akurasi terbaik dengan menggunakan KNN+GA. Pada ruang warna RGB didapatkan nilai akurasi sebesar 78.9% dan pada ruang warna HSV didapatkan nilai akurasi sebesar 75.8%. Terdapat peningkatan nilai akurasi yang dihasilkan oleh KNN+GA, hal tersebut membuktikan bahwa PSO berhasil mengoptimalkan hasil akurasi dari KNN dengan menentukan nilai k terbaiknya.

V. KESIMPULAN

Berdasarkan hasil pengujian serta analisis yang dilakukan, dapat disimpulkan bahwa GA dapat mengoptimasi KNN dengan memberikan hasil akurasi yang lebih tinggi jika dibandingkan dengan hanya menggunakan KNN. Hasil akurasi terbesar didapatkan dengan menggunakan ruang warna RGB dengan nilai akurasi sebesar 78.9% pada $k=9$ dan pada ruang warna HSV sebesar 75.8% dengan $k = 6$ atau 10. Hal tersebut menandakan bahwa nilai k yang sama, tidak dapat menghasilkan akurasi terbaik pada fitur yang berbeda sehingga perlu dioptimasi untuk mendapatkan nilai k terbaik yang dapat menghasilkan akurasi lebih tinggi untuk setiap fiturnya.

REFERENSI

- [1] R. Janandi and T. W. Cenggoro, "An Implementation of Convolutional Neural Network for Coffee Beans Quality Classification in a Mobile Information System," in *2020 International Conference on Information Management and Technology (ICIMTech)*, 2020, pp. 218–222.
- [2] F. Ribeyre and J. Avelino, "Impact of field pests and diseases on coffee quality," *Specialty coffee: managing coffee. International Plant Nutrition Institute. Malaysia: IPNI*, pp. 151–176, 2012.
- [3] J. G. M. Esgario, R. A. Krohling, and J. A. Ventura, "Deep learning for classification and severity estimation of coffee leaf biotic stress," *Computers and Electronics in Agriculture*, vol. 169, p. 105162, 2020.
- [4] E. M. T. Caballero and A. M. R. Duke, "Implementation of Artificial Neural Networks Using NVIDIA Digits and OpenCV for Coffee Rust Detection," in *2020 5th International Conference on Control and Robotics Engineering (ICCRE)*, 2020, pp. 246–251.
- [5] M. S. P. Rahul and M. Rajesh, "Image processing based Automatic Plant Disease Detection and Stem Cutting Robot," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020, pp. 889–894.
- [6] Y. P. Eko and others, "Bonferroni Mean Fuzzy K-Nearest Neighbors Based Handwritten Chinese Character Recognition," in *2021 International Conference on Data Science and Its Applications (ICoDSA)*, 2021, pp. 118–123.
- [7] Y. Herdiyeni and M. M. Santoni, "Combination of morphological, local binary pattern variance and color moments features for indonesian medicinal plants identification," in *2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2012, pp. 255–259.
- [8] N. Keen, "Color moments," *School of informatics, University of Edinburgh*, pp. 3–6, 2005.
- [9] B. T. Pham *et al.*, "Extreme learning machine based prediction of soil shear strength: a sensitivity analysis using Monte Carlo simulations and feature backward elimination," *Sustainability*, vol. 12, no. 6, p. 2339, 2020.
- [10] I. M. Zyout, "Toward automated detection and diagnosis of mammographic microcalcifications," 2010.
- [11] S. Mirjalili, "Genetic algorithm," in *Evolutionary algorithms and neural networks*, Springer, 2019, pp. 43–55.