

Klasifikasi Citra Penyakit Daun Kopi Arabika Menggunakan Support Vector Machine (SVM) Dengan Seleksi Fitur Information Gain

Sylviana Murni¹, Didit Widiyanto², Catur Nugrahaeni Puspita Dewi³

Program Studi Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

Email: sylvianam@upnvj.ac.id¹, didit.widiyanto@upnvj.ac.id², catur.nugrahaeni@upnvj.ac.id³

Abstrak. Tanaman kopi arabika lebih rentan terhadap penyakit, sehingga menyulitkan petani dalam memelihara tanaman kopi. Penelitian ini bermaksud untuk mempermudah proses identifikasi penyakit kopi arabika sejak dini dengan memanfaatkan teknik pengolahan citra digital. Tahapan yang dilakukan yaitu pra proses, ekstraksi fitur GLCM dan *Color Moments*, seleksi fitur *Information Gain*, dan klasifikasi *Support Vector Machine*. Proses ekstraksi fitur menghasilkan 29 fitur. Total data yang digunakan pada data *imbalanced* sebanyak 1325 data dan data *balanced* sebanyak 1534 data dengan pembagian 80% data latih dan 20% data uji. Dalam menentukan jumlah fitur terbaik yang memberikan akurasi tertinggi dilakukan pengujian performa klasifikasi dengan penggunaan jumlah fitur yang berbeda pada data *imbalanced* dan data *balanced*. Perbedaan jumlah fitur diperoleh dari kombinasi atribut peringkat teratas hingga terendah hasil seleksi fitur *Information Gain*. Hasil pengujian menunjukkan, penggunaan 18 fitur dari 29 fitur pada data *imbalanced* menghasilkan akurasi tertinggi dengan nilai akurasi sebesar 68,30%, presisi 55,77%, dan *recall* 57,85%.

Kata Kunci: Penyakit Daun, Tanaman Kopi Arabika, *Support Vector Machine* (SVM), *Information Gain*.

1. Pendahuluan

Kopi Arabika (*Coffea arabica L.*) adalah jenis kopi dengan nilai komersial tinggi dan merupakan produk ekspor utama Indonesia. Berdasarkan data *International Coffee Organization* (ICO), pada tahun 2019/2020 produksi kopi jenis arabika mengalami penurunan produksi sebesar 2.7%. Salah satu penyebab rendahnya produktivitas kopi disebabkan oleh Organisme Pengganggu Tanaman (OPT) berupa kelompok hama dan penyakit yang menyerang tanaman [1, pp. 1]. Tanaman kopi arabika lebih rentan terhadap penyakit dibandingkan dengan jenis tanaman kopi lainnya [2]. Hal tersebut menyulitkan petani kopi dalam memelihara tanaman kopi. Oleh sebab itu, petani perlu melakukan proses identifikasi penyakit sejak dini untuk mempercepat proses pengendalian penyakit yang dapat berpotensi meningkatkan jumlah produktivitas tanaman kopi. Untuk mempermudah proses identifikasi penyakit pada tanaman kopi arabika, maka diperlukannya proses identifikasi yang memanfaatkan teknik pengolahan citra digital dengan masukan berupa citra daun tanaman kopi arabika. Terdapat 3 tahapan dalam pengolahan citra digital, yaitu preprocessing, ekstraksi fitur, dan klasifikasi [3].

Penelitian mengenai ekstraksi fitur tekstur dan warna telah dilakukan sebelumnya oleh Tan et al dengan tujuan untuk menemukan metode ekstraksi fitur terbaik dengan membandingkan penerapan ekstraksi fitur tekstur dan warna dalam mengklasifikasikan citra penyakit daun tomat. Hasil akurasi dengan menerapkan ekstraksi fitur tekstur dan warna tersebut mendapatkan nilai akurasi tertinggi [4]. Pada penelitian kedua yang menerapkan kombinasi ekstraksi tekstur GLCM dan warna *Color Moment* HSV dapat mengenali 3 jenis penyakit pada daun kentang dan mendapatkan akurasi mencapai 87% [5].

Seleksi fitur merupakan teknik mereduksi fitur dengan memilih fitur-fitur penting dalam melakukan proses klasifikasi. Penelitian terkait seleksi fitur pada data citra membuktikan bahwa metode seleksi fitur *Information Gain* dapat memperoleh fitur-fitur relevan yang memberikan hasil akurasi tertinggi dibandingkan tanpa penggunaan *Information Gain* dimana penerapan 15 fitur terbaik mendapatkan akurasi sebesar 86,96% sedangkan penerapan seluruh fitur mendapatkan akurasi 60,86% [6]. Algoritma klasifikasi yang digunakan dalam penelitian ini ialah *Support Vector Machine* (SVM). Berdasarkan penelitian sebelumnya terkait dengan penggunaan algoritma SVM menyatakan bahwa penggunaan SVM mendapatkan hasil akurasi tertinggi dibandingkan dengan algoritma klasifikasi KNN, dan *Random Forest* [4].

Penelitian mengenai klasifikasi citra penyakit daun tanaman kopi arabika telah dilakukan sebelumnya oleh Esgario *et al* menggunakan algoritma klasifikasi *Convolutional Neural Network* (CNN) [7]. Berbeda dengan penelitian sebelumnya, penelitian ini mengusulkan klasifikasi citra penyakit daun kopi arabika menggunakan algoritma *Support Vector Machine* (SVM) dengan metode seleksi fitur *Information Gain* berdasarkan kelebihan-kelebihan hasil penelitian yang telah diuraikan. Tujuan penelitian ini untuk mengetahui tingkat akurasi dari metode klasifikasi *Support Vector Machine* (SVM) dalam mengklasifikasikan citra penyakit daun kopi arabika berdasarkan fitur-fitur yang relevan hasil proses seleksi fitur *Information Gain*.

2 Landasan Teori

2.1 Gray-level Co-Occurrence Matrix (GLCM)

GLCM merupakan metode untuk mendapatkan ekstraksi ciri tekstur orde kedua dengan menghitung probabilitas hubungan kedekatan antara dua buah piksel pada jarak (d) dan sudut tertentu (θ) [8]. Pada penelitian ini menggunakan 5 fitur dengan rumus sebagai berikut [9]:

1. Correlation

Correlation merepresentasikan ukuran korelasi linear antara nilai keabuan pada sebuah citra. Nilai *correlation* berada diantara -1 hingga 1. Berikut persamaan (1) untuk menghitung *Correlation*:

$$Correlation = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_{(i,j)}(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \quad (1)$$

2. Angular Second Moment (ASM)

Angular Second Moment (ASM) merupakan representasi ukuran homogenitas pada citra. Rumus *Angular Second Moment (ASM)* ditunjukkan pada persamaan (2) berikut:

$$ASM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_{(i,j)}^2 \quad (2)$$

3. Homogeneity

Homogeneity atau *Inverse Difference Moment (IDM)* mewakili ukuran keserbasamaan. Jika semua piksel memiliki nilai yang seragam, maka homogenitasnya akan tinggi. Nilai *Homogeneity* dapat dihitung melalui persamaan (3) berikut:

$$Homogeneity = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{p_{(i,j)}}{1+(i-j)^2} \quad (3)$$

4. Contrast

Contrast menyatakan perbedaan level keabuan pada piksel citra. Rumus *contrast* didefinisikan sebagai berikut:

$$Contrast = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_{(i,j)} |i - j|^2 \quad (4)$$

5. Entropy

Entropy merepresentasikan kompleksitas suatu citra. Semakin tinggi tingkat kompleksitas suatu citra maka semakin tinggi pula nilai entropinya. Berikut rumus untuk perhitungan nilai entropy:

$$Entropy = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_{(i,j)} \log(p_{(i,j)}) \quad (5)$$

Dimana:

μ_i = nilai rata – rata elemen baris

μ_j = nilai rata – rata elemen kolom

σ_i = nilai standar deviasi pada elemen baris matriks $p_{(i,j)}$

σ_j = nilai standar deviasi pada elemen kolom matriks $p_{(i,j)}$

2.1 Color Moment

Color Moment merupakan representasi dari entitas yang mencirikan karakteristik warna dari suatu citra berwarna. Informasi distribusi warna dihitung melalui perhitungan statistik pada masing-masing *channel* warna dengan menggunakan rumus sebagai berikut [10, pp. 313]:

a. *Mean*

Mean memberikan ukuran distribusi dengan rumus sebagai berikut:

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \quad (6)$$

b. *Standard Deviation*

Standard Deviation mewakili area sebaran distribusi. Akar kuadrat dari varians disebut standar deviasi. Rumus *Standard Deviation* adalah sebagai berikut:

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^2} \quad (7)$$

c. *Skewness*

Skewness atau kecondongan adalah ukuran ketidaksimetrisan. Berikut ini adalah rumus untuk menghitung *skewness*:

$$\theta = \frac{\sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^3}{MN\sigma^3} \quad (8)$$

Keterangan:

M = Panjang citra

N = Lebar citra

P_{ij} = nilai piksel pada baris ke i kolom j

MN = Jumlah piksel citra

2.2 Information Gain

Information Gain adalah metode seleksi fitur paling sederhana melalui penyortiran atribut, dan banyak diimplementasikan dalam aplikasi klasifikasi teks, analisis data gambar, dan analisis data microarray (Chormunge & Jena, dalam [11]). Langkah-langkah perhitungan *Information Gain* adalah sebagai berikut [6]:

1. Menghitung nilai entropi menggunakan persamaan (9). Entropi adalah ukuran ketidakpastian kelas yang memanfaatkan kemungkinan peristiwa atau atribut tertentu.

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (9)$$

2. Melakukan perhitungan *Information Gain* menggunakan rumus berikut:

$$Gain(S,A) = Entropy(S) - \sum_{values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (10)$$

Keterangan:

n = Jumlah nilai pada kelas klasifikasi

p_i = Jumlah sampel pada kelas i

A = Atribut

$Values(A)$ = himpunan nilai yang mungkin untuk A

$|S_v|$ = Jumlah sampel nilai v

$|S|$ = Total sampel data

$Entropy(S_v)$ = nilai *entropy* beberapa sampel dengan nilai v

2.3 Support Vector Machine (SVM)

Prinsip dasar SVM menggunakan prinsip linear *classifier*. Prinsip pengklasifikasian linear hanya dapat memisahkan dua kelas. Prinsip ini kemudian dikembangkan untuk menangani masalah nonlinear dengan menambahkan konsep kernel trick ke ruang kerja berdimensi lebih tinggi. Pengembangan kernel dapat dilakukan melalui urutan data dan grafik pada berbagai objek data (data diskrit dan data kontinu). Rumus dasar dari fungsi kernel dapat dinyatakan dalam persamaan (11) berikut:

$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (11)$$

Keterangan:

m = jumlah support vector

α_i = nilai bobot setiap data

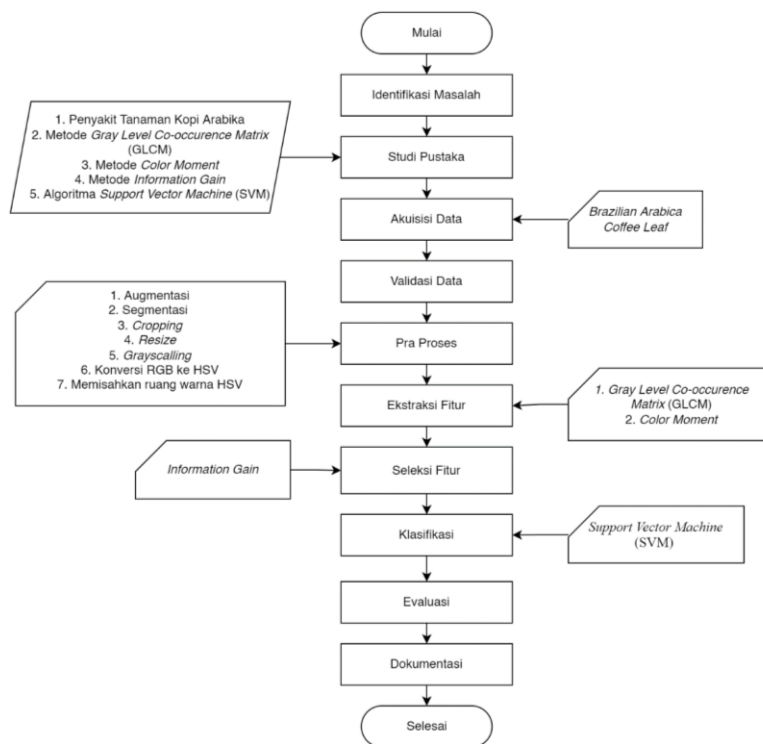
$K(x_i, x)$ = fungsi kernel

Syarat suatu fungsi menjadi fungsi kernel adalah memenuhi teorema Mercer bahwa matriks kernel yang dihasilkan harus semi-definite dan positif [12]. Pada penelitian ini fungsi kernel yang digunakan adalah *Radius Basis Function* (RBF) yang dinyatakan dalam persamaan berikut:

$$K(x_i, x) = \exp(-\gamma|x_i - x|^2), \gamma > 0 \quad (12)$$

3. Metode Penelitian

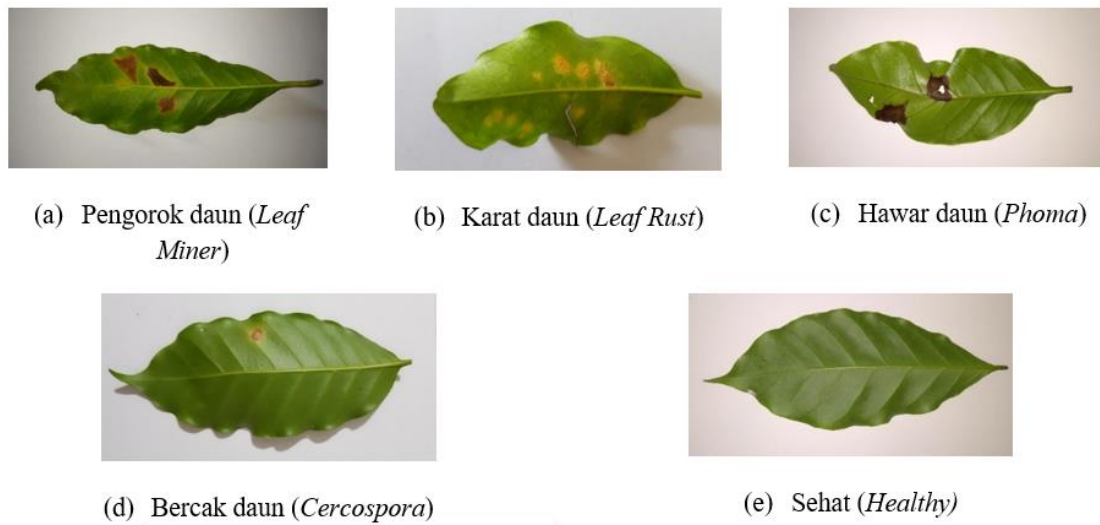
Dalam melakukan penelitian, untuk mencapai tujuan penelitian, maka peneliti membuat prosedur penelitian. Metodologi penelitian yang akan dilakukan dalam penelitian ini ditunjukkan pada Gambar. 1.



Gambar. 1. Metodologi Penelitian

3.1 Akuisisi Data

Akuisi data citra daun tanaman kopi arabika diperoleh dari situs Mendeley Data dengan judul “BRACOL - A Brazilian Arabica Coffee Leaf images dataset to identification and quantification of coffee diseases and pests” dalam situs <https://data.mendeley.com/datasets/yy2k5y8mxg/1> [13]. Dataset ini terdiri 5 kelas, yaitu pengorok daun (*Leaf Miner*), karat daun (*Leaf Rust*), hawar daun (*Phoma*), bercak daun (*Cercospora*), dan sehat (*Healthy*). Jumlah data sebanyak 1370 data citra berukuran 2048×1024 piksel. Dari 1370 data, 322 data merupakan data citra jenis penyakit pengorok daun (*Leaf Miner*), 348 data citra karat daun (*Leaf Rust*), 341 data citra hawar daun (*Phoma*), 87 data citra jenis penyakit bercak daun (*Cercospora*), dan 272 data citra daun kopi arabika sehat (*Healthy*).



Gambar. 2. 5 Kelas Citra Daun Tanaman Kopi Arabika Berpenyakit dan Sehat

3.2 Validasi Data

Validasi data dilakukan di Kantor Balai Pengkajian Teknologi Pertanian (BPTP) Jakarta dengan pakarnya adalah Ibu Dr. Ana Feronika Cindra Irawati, SP., MP. selaku peneliti tanaman dari BPTP Jakarta. Data citra yang telah divalidasi oleh pakar digunakan sebagai data citra input dalam proses penelitian ini. Pakar menyeleksi data citra pada masing-masing kelas berdasarkan pengamatannya. Perbandingan jumlah data sebelum dan sesudah dilakukan proses validasi oleh pakar dapat dideskripsikan pada Tabel 1 berikut:

Tabel 1. Jumlah Data Citra Daun Kopi Arabika

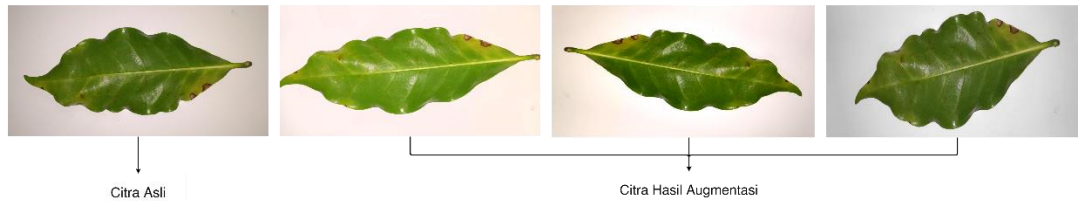
| No | Nama Kelas | Jumlah Data | |
|-------|-------------------------------------|------------------------|------------------------|
| | | Sebelum Validasi Pakar | Setelah Validasi Pakar |
| 1 | Pengorok daun (<i>Leaf Miner</i>) | 322 | 318 |
| 2 | Karat daun (<i>Leaf Rust</i>) | 348 | 324 |
| 3 | Hawar daun (<i>Phoma</i>) | 341 | 334 |
| 4 | Bercak daun (<i>Cercospora</i>) | 87 | 86 |
| 5 | Sehat (<i>Healthy</i>) | 272 | 263 |
| Total | | 1370 | 1325 |

3.3 Pra Proses

Pada tahap pra proses ini, akan dilakukan pengolahan data citra yang telah dilakukan proses validasi dengan tujuan untuk mendapatkan hasil data citra yang lebih mudah untuk diproses pada tahap ekstraksi ciri.

3.3.1 Augmentasi Data

Pada Tabel 1 diatas terlihat bahwa kelas Bercak daun (*Cercospora*) merupakan kelas dengan data minoritas sehingga perlu dilakukan *oversampling* menggunakan teknik augmentasi melakukan perubahan *brightness*, *saturation*, *contrast*, perbesaran, pergeseran, rotasi, dan membalikkan gambar horizontal dan vertikal secara acak. Hasil proses augmentasi pada kelas Bercak daun (*Cercospora*) didapatkan 209 data sehingga jumlah data pada kelas tersebut sebanyak 295 data dan total data secara keseluruhan sebesar 1534 data.



Gambar. 3. Citra Asli dan Citra Hasil Proses Augmentasi

3.3.2 Segmentasi Citra

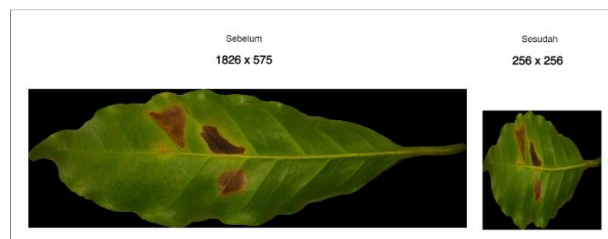
Metode segmentasi citra menggunakan *thresholding* dengan menentukan satu nilai ambang. Penggunaan nilai ambang sebesar 80 menghasilkan segmentasi citra biner yang optimal pada penelitian ini. Selanjutnya melakukan operasi morfologi *closing* untuk menutup lubang kecil pada area objek dan memperhalus. Kemudian citra hasil proses morfologi dikalikan dengan citra asli untuk mendapatkan citra RGB tersegmentasi.

3.3.3 Cropping

Bertujuan untuk memusatkan area objek dan mengurangi penggunaan *background*. Pada proses ini terdapat tahap *wrapping* dan *cropping*. *Wrapping* digunakan untuk mendeteksi tepi objek utama pada citra. Kemudian pembentukan *bounding box* berdasarkan hasil deteksi tepi dan melakukan pemotongan citra atau *cropping*.

3.3.4 Resize

Proses penyeragaman citra dilakukan agar tidak terjadi perbedaan ukuran citra dan dapat mengurangi beban komputasi. Proses *resize* dilakukan dengan mengubah ukuran citra menjadi menjadi 256 x 256 piksel dapat dilihat pada Gambar. 4 berikut.



Gambar. 4. Data Citra Sebelum dan Sesudah Proses *Resize*

3.3.5 Grayscale

Citra *grayscale* digunakan sebagai inputan untuk proses ekstraksi ciri tekstur GLCM. Pada tahap ini langkah yang perlu dilakukan adalah mengubah citra yang memiliki komponen warna yang terdiri dari 3 *channel* warna menjadi citra keabuan yang terdiri dari 1 *channel* warna.

3.3.6 Pemisahan Ruang Warna HSV

Proses pemisahan ruang warna HSV bertujuan untuk mendapatkan citra dari masing-masing *channel* H, S, dan V yang digunakan sebagai citra masukan proses ekstraksi ciri warna *Color Moment*.

3.4 Ekstraksi Fitur

3.4.1 Gray-Level Co-occurrence Matrix (GLCM)

Pada proses ini pengambilan nilai fitur menggunakan nilai jarak piksel sebesar 1 dengan 4 arah derajat yang berbeda yaitu 0°, 45°, 90°, dan 135° untuk menghasilkan matrix kookurensi. Kemudian menggunakan 5 variabel ciri tekstur GLCM antara lain *Correlation*, *Angular Second Moment (ASM)*, *Homogeneity*, *Contrast*, dan *Entropy*, sehingga menghasilkan 20 fitur yang menggambarkan ciri tekstur dari masing-masing citra. Beberapa nilai hasil ekstraksi fitur tekstur GLCM pada data *imbalanced* dideskripsikan pada Tabel 2 berikut:

Tabel 2. Hasil Ekstraksi Fitur GLCM Arah Sudut 0 Derajat

| Nama File | Contrast_0 | Homogeneity_0 | Entropy_0 | ASM_0 | Correlation_0 |
|------------------|------------|---------------|-----------|----------|---------------|
| phoma(331).jpg | 131.6799 | 0.528082 | 2.294358 | 0.174019 | 0.966179 |
| phoma(330).jpg | 85.2142 | 0.497321 | 2.389626 | 0.147997 | 0.974428 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| healthy(187).jpg | 50.31383 | 0.628203 | 2.003817 | 0.149777 | 0.979973 |
| healthy(207).jpg | 232.3055 | 0.488167 | 2.326085 | 0.166619 | 0.963376 |

3.4.2 Color Moment

Citra dari masing-masing *channel* warna *Hue (H)*, *Saturation (S)*, dan *Value (V)* melakukan perhitungan nilai dari *moment mean*, *standard deviation*, dan *skewness* sehingga menghasilkan 9 fitur warna yang terdiri dari *Mean_H*, *Std_H*, *Skewness_H*, *Mean_S*, *Std_S*, *Skewness_S*, *Mean_V*, *Std_V*, dan *Skewness_V*. Pada Tabel 3 menunjukkan beberapa nilai hasil proses ekstraksi fitur warna *Color Moment* pada data *imbalanced*.

Tabel 3. Hasil Ekstraksi Fitur Warna *Color Moment* Channel Warna H

| Nama File | Channel H | | | Channel S | | | Channel V | | |
|------------------|-----------|--------|----------|-----------|--------|----------|-----------|--------|----------|
| | Mean | Std | Skewness | Mean | Std | Skewness | Mean | Std | Skewness |
| phoma(331).jpg | 22.031 | 19.043 | -0.274 | 89.215 | 80.743 | -0.034 | 53.033 | 49.033 | 0.121 |
| phoma(330).jpg | 23.206 | 18.710 | -0.416 | 89.711 | 75.665 | -0.161 | 52.003 | 45.158 | 0.014 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| healthy(187).jpg | 25.664 | 20.652 | -0.432 | 93.389 | 76.706 | -0.312 | 48.820 | 40.924 | -0.202 |
| healthy(207).jpg | 22.813 | 19.291 | -0.328 | 71.279 | 64.429 | 0.035 | 71.739 | 61.862 | -0.215 |

3.5 Seleksi Fitur *Information Gain*

Data yang digunakan pada tahap ini adalah data hasil proses ekstraksi fitur dan data label kelas. Hasil dari seleksi fitur *Information Gain* adalah nilai gain dari masing-masing atribut. Kemudian hasil nilai gain tersebut akan diurutkan mulai dari atribut dengan nilai gain tertinggi hingga terendah untuk selanjutnya akan dilakukan klasifikasi. Urutan nilai dari hasil perhitungan seleksi fitur *Information Gain* ditunjukkan pada Tabel 4 dan Tabel 5 berikut:

Tabel 4. Hasil Seleksi Fitur *Information Gain* Data *Imbalanced*

| No | Fitur/Atribut | Nilai <i>Information Gain</i> | No | Fitur/Atribut | Nilai <i>Information Gain</i> |
|----|---------------|-------------------------------|----|---------------|-------------------------------|
| 1 | Std_S | 0.37709145 | 16 | Entropy_0 | 0.07890341 |
| 2 | Contrast_135 | 0.31608961 | 17 | Mean_V | 0.07627126 |
| 3 | Contrast_0 | 0.27810955 | 18 | ASM_135 | 0.07227294 |
| 4 | Contrast_45 | 0.26009196 | 19 | ASM_0 | 0.0697233 |
| 5 | Contrast_90 | 0.25145072 | 20 | Skewness_V | 0.05988585 |
| 6 | Mean_H | 0.22776004 | 21 | ASM_90 | 0.05413536 |

| No | Fitur/Atribut | Nilai Information Gain | No | Fitur/Atribut | Nilai Information Gain |
|----|-----------------|------------------------|----|-----------------|------------------------|
| 7 | Std_H | 0.21620064 | 22 | ASM_45 | 0.04829465 |
| 8 | Std_V | 0.21523335 | 23 | Homogeneity_45 | 0.04542555 |
| 9 | Mean_S | 0.18108086 | 24 | Homogeneity_0 | 0.0445811 |
| 10 | Skewness_H | 0.1489849 | 25 | Entropy_45 | 0.03820278 |
| 11 | Correlation_0 | 0.13836264 | 26 | Homogeneity_90 | 0.03583725 |
| 12 | Correlation_135 | 0.13731413 | 27 | Entropy_135 | 0.0330995 |
| 13 | Correlation_45 | 0.13216659 | 28 | Entropy_90 | 0.03145669 |
| 14 | Skewness_S | 0.09149689 | 29 | Homogeneity_135 | 0.02575749 |
| 15 | Correlation_90 | 0.0907112 | | | |

Berdasarkan Tabel 4 urutan fitur yang relevan pada data *imbalanced* berdasarkan hasil seleksi fitur *Information Gain* yaitu Std_S, Contrast_135, Contrast_0, Contrast_45, Contrast_90, Mean_H dan seterusnya hingga fitur terakhir yaitu Homogeneity_135.

Tabel 5. Hasil Seleksi Fitur Information Gain Data *Balanced*

| No | Fitur/Atribut | Nilai Information Gain | No | Fitur/Atribut | Nilai Information Gain |
|----|-----------------|------------------------|----|-----------------|------------------------|
| 1 | Std_S | 0.31213581 | 16 | Skewness_S | 0.11154097 |
| 2 | Contrast_135 | 0.26058904 | 17 | ASM_45 | 0.096309 |
| 3 | Std_V | 0.25602526 | 18 | ASM_135 | 0.09269355 |
| 4 | Contrast_90 | 0.23782785 | 19 | ASM_90 | 0.09211567 |
| 5 | Mean_H | 0.22933856 | 20 | ASM_0 | 0.09115507 |
| 6 | Contrast_0 | 0.21987663 | 21 | Correlation_90 | 0.09080093 |
| 7 | Contrast_45 | 0.21657481 | 22 | Entropy_0 | 0.07206645 |
| 8 | Std_H | 0.16643869 | 23 | Entropy_135 | 0.06954105 |
| 9 | Correlation_0 | 0.15274448 | 24 | Entropy_90 | 0.06757643 |
| 10 | Correlation_45 | 0.13323258 | 25 | Entropy_45 | 0.05897294 |
| 11 | Mean_S | 0.13149087 | 26 | Homogeneity_90 | 0.04791138 |
| 12 | Skewness_V | 0.12838066 | 27 | Homogeneity_135 | 0.04288276 |
| 13 | Correlation_135 | 0.12729334 | 28 | Homogeneity_0 | 0.04198779 |
| 14 | Skewness_H | 0.12259653 | 29 | Homogeneity_45 | 0.01944827 |
| 15 | Mean_V | 0.11578129 | | | |

Berdasarkan Tabel 5 urutan fitur yang relevan pada data *balanced* berdasarkan hasil seleksi fitur *Information Gain* yaitu Std_S, Contrast_135, Std_V, Contrast_90, Mean_H, Contrast_0 dan seterusnya hingga fitur terakhir yaitu Homogeneity_45.

3.6 Pembagian Data

Data hasil ekstraksi ciri pada data *imbalanced* yang terdiri dari 1325 baris dan 29 fitur, serta data hasil ekstraksi ciri pada data *balanced* yang terdiri dari 1534 baris dan 29 fitur dilakukan pembagian data menggunakan perbandingan pembagian data latih dan data uji sebesar 80:20 secara acak dan seimbang. Pada data *imbalanced* terdapat 1060 data latih dan 265 data uji, sedangkan pada data *balanced* terdapat 1227 data latih dan 307 data uji.

3.3 Klasifikasi

Penerapan algoritma SVM pada penelitian ini menggunakan fungsi SVC dari *library scikit learn* dengan menggunakan kernel *Radial Basis Function* (RBF), dan parameter gamma dengan nilai *scale*, C dengan nilai 100 dan 1000, serta *decision function* dengan nilai *ovr* yang dapat dilihat pada *syntax* berikut:

```
SVM_model = svm.SVC(kernel='rbf', gamma="scale", C=100, decision_function_shape='ovr')
```


Keterangan :

svm.SVC = fungsi model klasifikasi SVM.

kernel = pengaturan jenis kernel trick yang akan digunakan {'linear', 'poly', 'sigmoid', 'precomputed', 'rbf'}.

C = cost atau parameter penentu besaran penalti akibat kesalahan klasifikasi.

gamma = {'scale', 'auto'}.

decision_function_shape = bentuk fungsi keputusan {'ovo', 'ovr'}.

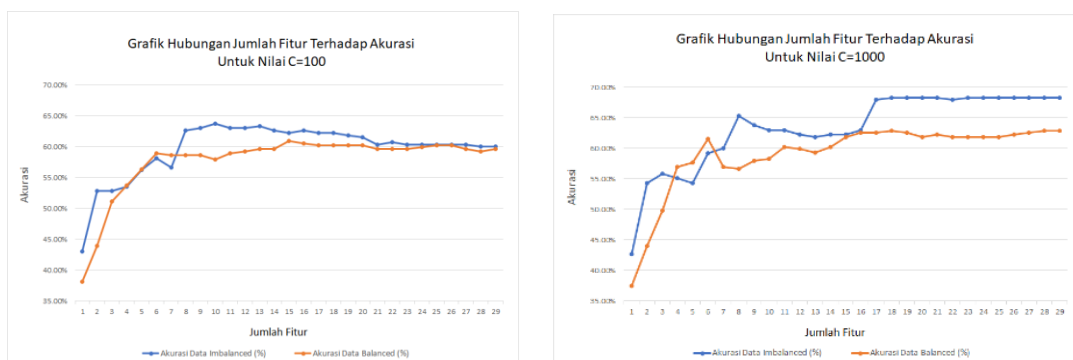
4. Hasil dan Pembahasan

Pada tahap ini akan membahas hasil pengujian performa klasifikasi dengan penggunaan jumlah atribut yang berbeda pada data *imbalanced* dan data *balanced* dan nilai parameter C (*Cost*) pada SVM yang berbeda yaitu C=100 dan C=1000. Variasi jumlah atribut diperoleh dari kombinasi atribut peringkat teratas hingga terendah hasil seleksi fitur *Information Gain* pada Tabel 4 dan Tabel 5. Berdasarkan hasil pengujian, jumlah kombinasi fitur terbaik dengan nilai akurasi tertinggi dapat diringkas dalam bentuk Tabel 6 berikut ini:

Tabel 6. Hasil Pengujian Jumlah Fitur Terbaik

| Keseimbangan Data | Nilai Parameter C | Jumlah Kombinasi Fitur | Akurasi (%) | Presisi (%) | Recall (%) |
|------------------------|-------------------|------------------------|-------------|-------------|------------|
| Data <i>Imbalanced</i> | C=100 | 10 | 63,77 % | 51,91 % | 54,18 % |
| | C=1000 | 18 | 68,30 % | 55,77 % | 57,85 % |
| Data <i>Balanced</i> | C=100 | 15 | 60,91 % | 64,38 % | 59,95 % |
| | C=1000 | 18 | 62,86 % | 64,10 % | 61,71 % |

Berdasarkan pada hasil pengujian jumlah fitur terbaik pada Tabel 6 diatas, penggunaan kombinasi 18 jumlah fitur pada klasifikasi citra penyakit kopi arabika menggunakan data *imbalanced* dan nilai parameter C=1000 mendapatkan hasil terbaik dengan nilai akurasi sebesar 68,30%, presisi 55,77%, dan *recall* 57,85%. 18 fitur terbaik tersebut, yaitu Std_S, Contrast_135, Contrast_0, Contrast_45, Contrast_90, Mean_H, Std_H, Std_V, Mean_S, Skewness_H, Correlation_0, Correlation_135, Correlation_45, Skewness_S, Correlation_90, Entropy_0, Mean_V, ASM_135.



Gambar. 5. Grafik Hubungan Jumlah Fitur Terhadap Akurasi Untuk Nilai C (*Cost*) =100 dan C (*Cost*) =1000

Pada grafik Gambar 5 terlihat bahwa penggunaan jumlah fitur yang berbeda pada nilai C=100 dan C=1000 mempengaruhi akurasi dengan hasil yang tidak stabil atau mengalami kenaikan dan penurunan. Pada penggunaan nilai C=100 nilai akurasi tertinggi diperoleh saat menggunakan 10 fitur pada data *imbalanced* dan 15 fitur pada data *balanced*, sedangkan pada penggunaan nilai C=1000 nilai akurasi tertinggi diperoleh saat menggunakan 18 fitur pada data *imbalanced* dan data *balanced*. Penggunaan jumlah fitur yang terlalu sedikit menghasilkan nilai akurasi klasifikasi yang buruk. Pada nilai C=100 penggunaan seluruh fitur tidak memberikan akurasi yang maksimal dan pada nilai C=1000 penggunaan seluruh fitur menghasilkan nilai akurasi yang sama dengan penggunaan 18 fitur.

5. Penutup

5.1 Kesimpulan

Dari 29 fitur hasil ekstraksi fitur GLCM dan *Color Moment*, nilai akurasi tertinggi dalam mengklasifikasikan citra penyakit daun kopi arabika menggunakan algoritma *Support Vector Machine* (SVM) dihasilkan pada penggunaan 18 fitur menggunakan data *imbalanced* dan nilai parameter $C=1000$. Hasil evaluasi diperoleh nilai akurasi sebesar 68,30%, presisi sebesar 55,77%, dan *recall* sebesar 57,85%. Metode seleksi fitur *Information Gain* dapat menentukan jumlah fitur terbaik dalam melakukan klasifikasi citra penyakit daun kopi arabika. Hasil akurasi tertinggi diperoleh pada penggunaan 18 fitur terbaik hasil seleksi fitur *Information Gain*, yaitu Std_S, Contrast_135, Contrast_0, Contrast_45, Contrast_90, Mean_H, Std_H, Std_V, Mean_S, Skewness_H, Correlation_0, Correlation_135, Correlation_45, Skewness_S, Correlation_90, Entropy_0, Mean_V, ASM_135.

5.2 Saran

Saran yang diberikan untuk pengembangan penelitian selanjutnya, yaitu menambahkan metode ekstraksi ciri bentuk atau metode ekstraksi ciri tekstur dan warna lainnya untuk meningkatkan jumlah variasi fitur yang akan digunakan dalam proses seleksi fitur, menggunakan metode seleksi fitur lainnya dalam menemukan fitur yang relevan seperti F-Score, *Correlation based Feature Selection* (CFS) untuk mengetahui metode seleksi fitur terbaik, serta menggunakan algoritma klasifikasi lainnya dan melakukan perbandingan dengan algoritma klasifikasi SVM agar mendapatkan hasil klasifikasi yang lebih baik.

Referensi

- [1] R. Harni *et al.*, *Teknologi Pengendalian Hama dan Penyakit Tanaman Kopi*. 2015.
- [2] L. Sugiarti, "Analisis Tingkat Keparahan Penyakit Karat Daun Pada Tanaman Kopi Arabika Di Kebun Percobaan Fakultas Pertanian Universitas Winaya Mukti Tanjungsari," *Jagros : Jurnal Agroteknologi dan Sains (Journal of Agrotechnology Science)*, vol. 1, no. 2. p. 80, 2017, doi: 10.52434/jagros.v1i2.309.
- [3] M. R. Rasyid, Z. Tahir, and N. Syafaruddin, "Pengolahan Citra Digital untuk Mendeteksi Kesalahan Kerja Mesin Industri dengan Metode Learning Vector Quantization," *J. Pekommas*, vol. 4, no. 2, p. 131, 2019, doi: 10.30818/jpkm.2019.2040203.
- [4] L. Tan, J. Lu, and H. Jiang, "Tomato Leaf Diseases Classification Based on Leaf Images: A Comparison between Classical Machine Learning and Deep Learning Methods," *AgriEngineering*, vol. 3, no. 3, pp. 542–558, 2021, doi: 10.3390/agriengineering3030035.
- [5] P. U. Rakhmawati, Y. M. Pranoto, and E. Setyati, "Klasifikasi Penyakit Daun Kentang Berdasarkan Fitur Tekstur Dan Fitur Warna Menggunakan Support Vector Machine," *Semin. Nas. Teknol. dan Rekayasa 2018*, no. 4, pp. 1–8, 2018, [Online]. Available: <http://research-report.umm.ac.id/index.php/sentra/article/view/2127>.
- [6] V. W. Saputra, Y. A. Sari, and A. W. Widodo, "Klasifikasi Jenis Makanan menggunakan Neighbor Weighted K-Nearest Neighbor dengan Seleksi Fitur Information Gain," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. Mei 2019, pp. 4749–4758, 2019.
- [7] J. G. M. Esgario, R. A. Krohling, and J. A. Ventura, "Deep learning for classification and severity estimation of coffee leaf biotic stress," *Comput. Electron. Agric.*, vol. 169, pp. 1–11, 2019, doi: 10.1016/j.compag.2019.105162.
- [8] F. Z. Rahmanti, N. K. Ningrum, S. E. Sukmana, and P. W. Adi, "Plasmodium Falciparum Identification in Thick Blood Preparations Using GLCM and Support Vector Machine (SVM)," *J. Appl. Intell. Syst.*, vol. 2, no. 1, pp. 12–20, 2017, doi: 10.33633/jais.v2i1.1388.
- [9] E. Alvansa, "Pengenalan Tekstur Menggunakan Metode Glcm Serta Modul Nirkabel," 2019.
- [10] A. Kadir and A. Susanto, *Teori dan Aplikasi Pengolahan Citra*. Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia: ANDI, 2013.
- [11] D. S. Atmaja, Y. A. Sari, and R. C. Wihandika, "Seleksi Fitur Information Gain pada Klasifikasi Citra Makanan Menggunakan Ekstraksi Fitur Haralick dan YUV Color Moment," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 2, pp. 1917–1924, 2019.
- [12] D. T. E. Situmorang, "Klasifikasi Artikel Berita Berdasarkan UU ITE No.11 2008 Menggunakan Support Vector Machine," *Repos. Institusi USU*, no. 11, 2021.
- [13] Krohling, Renato; esgario, José; Ventura, Jose A. (2019), "BRACOL - A Brazilian Arabica Coffee Leaf images dataset to identification and quantification of coffee diseases and pests", *Mendeley Data*, V1, doi: 10.17632/yy2k5y8mxg.1.