

Penerapan *Random Forest* untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel

Luthfiana Ratnawati dan Dwi Ratna Sulistyaningrum
Departemen Matematika, Fakultas Matematika Komputasi dan Sains Data,
Institut Teknologi Sepuluh Nopember (ITS)
e-mail: dwiratna@gmail.com

Abstrak—Tingkat keparahan penyakit pada tanaman merupakan salah faktor penting untuk diketahui sebagai upaya pengendalian hama dan penyakit yang dapat berpengaruh penting dalam perkembangan tanaman. Teknologi pengolahan citra digital (*Digital Image Processing*) saat ini berkembang semakin pesat, salah satunya dalam bidang pertanian. Pada penelitian tugas akhir ini, dilakukan pengukuran tingkat keparahan penyakit pada daun apel dengan menggunakan metode klasifikasi *Random Forest*. Pengukuran tingkat keparahan penyakit pada daun apel dilakukan dalam beberapa tahapan proses yaitu pra-pengolahan citra, segmentasi citra menggunakan *K-means clustering*, ekstraksi fitur ukuran, bentuk dan warna pada citra dan yang terakhir klasifikasi menggunakan metode *Random Forest*. Data citra yang digunakan sejumlah 467 citra daun apel dan menghasilkan kinerja klasifikasi *Random Forest* dengan akurasi yang menunjukkan bahwa metode *Random Forest* mampu mengukur tingkat keparahan penyakit pada daun apel dengan akurasi tertinggi pada proses pelatihan sebesar 100% dan nilai akurasi tertinggi pada proses pengujian sebesar 75.3191%.

Kata Kunci—Penyakit Tanaman, Daun Apel, Tingkat Keparahan, *Random Forest*.

I. PENDAHULUAN

INDONESIA yang memiliki iklim tropis merupakan tempat yang baik untuk tumbuh beberapa tanaman subtropika. Salah satu tanaman subtropika yang dapat tumbuh di Indonesia selain jeruk, anggur dan stroberi adalah apel. Pada budidaya tanaman apel, pengendalian hama dan penyakit merupakan salah satu faktor penting dalam perkembangan tanaman apel karena dapat mempengaruhi hasil panen buah apel. Seperti yang disebutkan dalam data Badan Pusat Statistik (BPS) Kota Batu pada tahun 2017 bahwa jumlah pohon menghasilkan dan produksi buah apel adalah 972.845 tanaman menghasilkan 145.927 Kg buah apel pada Triwulan I, 970.863 tanaman menghasilkan 145.633 Kg buah apel pada Triwulan II dan semakin menurun menjadi 886.850 tanaman menghasilkan 133.028 Kg buah apel pada Triwulan III. Beberapa kasus yang menyebabkan penurunan jumlah tanaman yang menghasilkan dan produksi buah apel selain iklim dan semakin banyaknya tanaman berumur tua, juga dikarenakan tanaman apel terserang hama dan penyakit. Oleh karena itu, diperlukan identifikasi serta penanganan sedini mungkin agar mengurangi ancaman semakin menurunnya jumlah produksi buah apel.

Penanganan terhadap penyakit pada tanaman apel dapat dilakukan dengan beberapa cara, seperti mengatur jarak tanam yang tidak terlalu rapat, pemusnahan tanaman atau bagian yang terinfeksi penyakit dan penyemprotan fungisida [1]. Penanganan dan pendeteksian penyakit tanaman ini

tentunya menghabiskan biaya yang banyak jika dilakukan pada lahan perkebunan yang luas. Penggunaan fungisida yang berlebihan untuk pengobatan penyakit tanaman dapat meningkatkan bahaya tingkat residu racun dan dapat mengakibatkan pencemaran lingkungan. Oleh karena itu perlu diidentifikasi tingkat keparahan suatu penyakit pada tanaman agar dapat meminimalkan biaya yang dikeluarkan serta mengurangi tingkat pencemaran lingkungan. Identifikasi penyakit tanaman sendiri juga memerlukan kemampuan dan ketelitian yang tinggi mengingat metode yang digunakan masih menggunakan metode manual, yakni dengan mata telanjang. Metode ini sebenarnya juga bisa digunakan untuk menentukan tingkat keparahan penyakit pada tanaman. Namun, dalam praktiknya hasil yang diperoleh masih subjektif dan kurang tepat serta memakan waktu yang banyak.

Teknologi pengolahan citra digital (*Digital Image Processing*) saat ini berkembang semakin pesat dengan adanya berbagai aplikasi dalam berbagai bidang, seperti kedokteran, industri, pertanian, geologi, kelautan dan lain sebagainya. Penerapan yang semakin luas dari teknologi ini mendorong penelitian yang bergerak pada bidang pertanian dan perkebunan untuk juga memanfaatkannya sebagai upaya mempermudah beberapa hal di dalamnya, seperti pendeteksian penyakit pada tanaman.

Pengolahan citra digital mempunyai beberapa proses agar dapat diperoleh informasi yang diperlukan dari input berupa citra tersebut. Salah satu proses penting dalam pengolahan citra adalah segmentasi yang bertujuan untuk memisahkan wilayah objek dengan latar belakang. Metode segmentasi yang umum digunakan salah satunya adalah dengan menerapkan algoritma K-means yang juga diterapkan dalam penelitian oleh K. S. Archana dan Arun Sahayadhas yang berjudul “Automatic Rice Leaf Disease Segmentation Using Image Processing Techniques” pada tahun 2018 [2]. Selain itu, proses penting lainnya yang dapat membantu dalam pengolahan citra adalah proses klasifikasi agar lebih mudah diambil kesimpulan dan informasi yang dibawa citra tersebut. Metode *Random Forest* yang merupakan modifikasi dari *Decision Pohon keputusan* juga telah banyak diterapkan dalam klasifikasi citra. Seperti penelitian pada tahun 2015 yang berjudul “Random Forest Based Classification of Medical X-Ray Images Using a Genetic Algorithm For Feature Selection” oleh Imane Nedjar, Mostafa El Habib Dahou dan Nesma Settouti [3].

Oleh karena itu penelitian ini diadakan untuk menerapkan metode *Random Forest* untuk mengukur tingkat keparahan penyakit pada daun apel. Dari hasil penelitian yang dilakukan nantinya diharapkan dapat membantu mempermudah pekerjaan dalam penanganan dan identifikasi penyakit pada daun apel.

Sehingga, dapat dilakukan tindakan lanjutan untuk mencegah penyakit semakin bertambah parah dan tidak menyebar ke tanaman apel lainnya, yang kemudian dapat mempengaruhi penurunan produksi buah apel.

II. TINJAUAN PUSTAKA

A. Pengolahan Citra Digital

Pengolahan Citra Digital adalah ilmu yang mempelajari hal-hal yang berkaitan dengan perbaikan kualitas gambar (peningkatan kontras, transformasi warna, restorasi citra), transformasi gambar (rotasi, translasi, skala, tranformasi geometrik), melakukan pemilihan citra ciri (*feature images*) yang optimal untuk tujuan analisis, melakukan proses penarikan informasi atau deskripsi obyek yang terkandung pada citra, melakukan kompresi atau reduksi data untuk tujuan penyimpanan data, transmisi data, dan waktu proses data. Masukan dari pengolahan citra adalah citra, sedangkan outputnya adalah citra hasil pengolahan. Pengolahan citra adalah pemrosesan citra, dengan maksud untuk mendapatkan kualitas citra yang diinginkan [4].

Pengolahan citra dikelompokkan menjadi dua yaitu memperbaiki citra sesuai dengan kebutuhan dan mengolah informasi yang terdapat di dalam citra. Mengolah informasi dalam citra umumnya untuk mengolah objek citra dengan cara mengekstraksi informasi penting yang ada di dalamnya. Dengan pengolahan citra diharapkan citra yang diproses dapat diambil cirinya [5].

B. Segmentasi K-Means

Segmentasi citra adalah proses pengolahan citra yang bertujuan memisahkan wilayah (*region*) objek dengan wilayah latar belakang agar objek mudah dianalisis dalam rangka mengenali objek yang banyak melibatkan persepsi visual.

Proses segmentasi citra didasarkan pada perbedaan derajat keabuan citra. Untuk mengubah citra berwarna yang mempunyai nilai matrik masing-masing r, g dan b menjadi citra *grayscale* dengan nilai s, maka dilakukan konversi dengan mengambil rata-rata dari nilai r, g dan b [6].

$$s = \frac{r+g+b}{3} \tag{1}$$

Secara sederhana *K-means* adalah algoritma untuk melakukan klasifikasi berdasarkan atribut ke dalam *K* grup. Dimana *K* adalah bilangan bulat positif. Pengelompokan dilakukan dengan meminimalisir jarak antar data dan berdasarkan pengelompokan terpusat [7]. Jarak yang digunakan dalam *K-means clustering* adalah jarak Euclidean (*Euclidean Distance*). Jarak Eiclidean adalah jarak antara dua titik yang terdapat dalam ruang dimensi *n* dengan menghitung akar perbedaan antara dua titik atau objek yang berpasangan.

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{2}$$

dengan *x* = titik data pertama
y = titik data kedua
n =ruang dimensi
i = 1,2, ..., *n*

Algoritme K-means pada dasarnya terdiri dari tiga tahap, yaitu :

1. Inisialisasi : Pilih *K*, yang merupakan set awal *centroid*, yaitu titik virtual pada ruang data yang dibuat acak.

2. Setiap titik dari dataset ditetapkan pada *centroid* terdekatnya.
3. Posisi dari *centroid* diperbarui berdasarkan rata-rata dari titik data yang telah ditetapkan ke kelompok tersebut. Dengan kata lain, *centroid* tersebut bergerak menuju tengah dari titik yang telah ditetapkan.

Algoritma *K-means* akan menemukan *K* kelompok dari data yang ditunjukkan pada fungsi objektif sebagai berikut [7]:

$$F = \sum_{j=1}^N \sum_{i=1}^K a_{ji} D(x_j, C_i) \tag{3}$$

dengan *C_i* dan *a_{ji}* diperoleh melalui persamaan sebagai berikut:

$$C_i = \frac{1}{N} \sum_{j=1}^N x_j \tag{4}$$

$$a_{ji} = \begin{cases} 1, & d = \min(D(x_j, C_i)) \\ 0, & \text{lainnya} \end{cases} \tag{5}$$

dimana *N* = jumlah data

K = jumlah *cluster*

a_{ji} = alokasi keanggotaan titik

D = jarak *Euclidean* (2)

C_i = *centroid*

x_j = nilai data

i = 1,2, ..., *K*

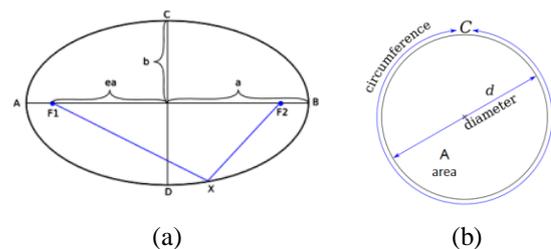
j = 1,2, ..., *N*

Dalam penggunaan algoritme *K-means* dalam segmentasi citra, untuk setiap piksel dari suatu gambar dikaitkan dengan warna yang dijelaskan dalam RGB. Gambar yang akan disegmentasi kemudian dapat direpresentasikan sebagai seperangkat titik dalam ruang data 3D.

C. Ekstraksi Fitur

Tujuan utama ekstraksi fitur adalah untuk memperoleh informasi yang paling relevan dari data asli dan merepresentasikan informasi tersebut dalam ruang dimensi yang lebih rendah [8]. Fitur yang sering diekstrak antara lain fitur bentuk, fitur geometri, fitur ukuran, fitur tekstur dan fitur warna. Dimana setiap fitur yang diekstrak mempunyai perbedaan parameter untuk mendapatkan fitur tersebut.

Fitur ukuran mempunyai beberapa parameter seperti *area* atau luas yang merupakan banyaknya piksel yang menyusun suatu objek. Sedangkan *perimeter* atau keliling adalah banyaknya piksel yang berada pada *boundary* objek.



Gambar 1. Ilustrasi Parameter fitur bentuk (a) Eccentricity (b) Perimeter.

Fitur Bentuk bertujuan untuk membedakan bentuk objek satu dengan objek lainnya, dapat menggunakan parameter yang disebut dengan '*eccentricity*'. *Eccentricity* merupakan nilai perbandingan antara jarak *foci ellips minor* dengan *foci ellips mayor* suatu objek. *Eccentricity* memiliki rentang nilai antara 0 hingga 1. Terdapat pula parameter lainnya yang dapat digunakan untuk membedakan bentuk suatu objek yaitu '*metric*'. *Metric* merupakan nilai perbandingan antara

luas dan keliling objek. *Metric* memiliki rentang nilai antara 0 hingga 1. Ilustrasi *eccentricity* dan *metric* diilustrasikan pada Gambar 1.

Sedangkan fungsi yang digunakan untuk menghitung nilai *eccentricity* dan *metric* adalah sebagai berikut :

a. *Eccentricity*

$$e = \sqrt{1 - \frac{b^2}{a^2}} \tag{6}$$

b. *Metric*

$$M = \frac{4\pi \times Area}{Perimeter^2} \tag{7}$$

dengan a = *axis minor*

b = *axis mayor*

Area = luas objek

Perimeter = keliling objek

Fitur lain dari citra adalah fitur warna yang terdiri dari *mean*, standart deviasi, *kurtosis* dan *skewness* dengan penjelasan serta fungsi perhitungannya sebagai berikut [9]:

a. *Mean*

Mean merupakan rata-rata nilai piksel (P_{ij}) pada masing-masing *channel* warna pada citra.

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \tag{8}$$

b. Standart Deviasi

$$\sigma = \sqrt{\frac{1}{MN-1} \sum_{i=1}^M \sum_{j=1}^N |P_{ij} - \mu|^2} \tag{9}$$

dengan M dan N secara berturut-turut adalah nilai panjang dan lebar piksel pada citra dan P_{ij} adalah nilai piksel pada citra.

c. *Skewness* dan *Kurtosis*

Skewness adalah derajat ketidaksimetrisan suatu distribusi. Apabila *skewness* = 0 maka dikatakan simetris. *Kurtosis* adalah derajat keruncingan suatu distribusi (biasanya diukur relatif terhadap distribusi normal). Distribusi yang dimaksud adalah warna. Secara perhitungan, *skewness* adalah momen ketiga terhadap mean dan *kurtosis* adalah momen keempat terhadap mean. *Skewness* dan *kurtosis* merupakan rata-rata nilai piksel (P_{ij}) dikurangi mean (μ) kemudian dipangkatkan dengan n pada masing-masing *channel* dari citra.

$$\theta = \frac{\sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^n}{MN\sigma^n} \tag{10}$$

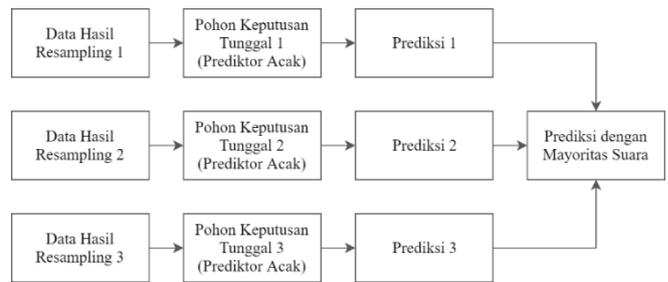
Dimana untuk *skewness* nilai n adalah 3, dan untuk *kurtosis* nilai n adalah 4.

D. *Random Forest*

Random forest merupakan metode klasifikasi yang dilakukan dengan mengembangkan metode *Decision Tree* berdasarkan pemilihan atribut acak pada setiap *node* untuk menentukan klasifikasi. Pada proses klasifikasinya didasarkan pada suara terbanyak dari pohon keputusan yang dikembalikan.

Random forest dapat dibangun dengan menggunakan *bagging* dengan pemilihan atribut acak. Metode CART (*Classification and Regression Tree*) digunakan untuk menumbuhkan pohon keputusan. Pohon keputusan tersebut tumbuh hingga ukuran maksimum dan tidak akan dipangkas. Sehingga dihasilkanlah kumpulan pohon yang kemudian disebut hutan (*forest*) [10].

Random Forest adalah suatu metode klasifikasi yang terdiri dari kumpulan terstruktur pohon keputusan dimana vektor acak independen didistribusikan secara identik dan setiap pohon keputusan memberikan suara unit untuk kelas paling populer pada masukan x , seperti pada Gambar 2 [11].



Gambar 2. Algoritma sederhana *Random Forest*.

Random Forest memiliki kelebihan sebagai berikut [11]:

1. Akurasi bagus.
2. Relatif kuat terhadap outliers dan noise.
3. Lebih cepat dari pada *Bagging* dan *Boosting*.
4. Sederhana dan mudah diparalelkan.

Proses untuk membangun *Random Forest* adalah sebagai berikut [11]:

1. Ketika sampel *bootstrap* dibangun dengan mengambil sampel data dengan penggantian setiap pohon keputusan, maka sepertiga dari contoh ditinggalkan.
2. Contoh yang ditinggalkan dikenal sebagai data OOB (*Out of Bag*).
3. Setiap pohon keputusan pada *forest* memiliki data OOB sendiri yang digunakan untuk estimasi kesalahan masing-masing pohon keputusan yang dikenal sebagai estimasi kesalahan OOB.
4. *Random Forest* juga dapat menghitung tingkat kepentingan dan perkiraan variabel. Perkiraan digunakan untuk menghapus dan mengganti nilai-nilai dan *outlier* yang hilang.

E. Pengukuran Performa

Pengukuran performa klasifikasi dapat dilihat berdasarkan nilai *recall*, *precision* dan akurasi. *Precision* berfungsi untuk mengukur ketepatan pengklasifikasi. Sedangkan *recall* untuk mengukur kelengkapan pengklasifikasi. Akurasi berfungsi untuk mengukur keseluruhan performa pengklasifikasi [12].

Nilai akurasi, *recall* dan *precision* diperoleh melalui perhitungan berdasarkan *matriks confusion* seperti pada Tabel 1.

Tabel 1.

Tabel *matrix confusion*

Matrix Confussion	Prediksi				
	0	1	2	3	
Aktual	0	TP	FN	FN	FN
	1	FP	TN	FN	FN
	2	FP	FN	TN	FN
	3	FP	FN	FN	TN

Sebagai contoh untuk klasifikasi ke dalam empat kelas pada Tabel 1., terdapat *matrix confusion* dengan kelas 0 sebagai patokan, maka TP (*True Positive*) adalah pada pemetaan kelas aktual 0 ke kelas prediksi 0. FP (*False Positive*) adalah pemetaan kelas aktual non 0 ke kelas prediksi 0, TN (*True Negative*) adalah pemetaan kelas aktual non 0 yang benar diprediksikan sedangkan FN (*False*

Negative) adalah pemetaan kelas aktual non 0 ke kelas prediksi non 0 yang salah.

Perhitungan nilai recall, precision dan akurasi berdasarkan pada formula sebagai berikut [12]:

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \tag{10}$$

Begitu pula berlaku untuk kelas ke 1, 2 dan 3 dengan penetapan TP, TN, FP dan FN sesuai definisi dan perhitungan *recall*, *precision* dan akurasi berturut-turut menggunakan persamaan 8, 9 dan 10.

F. Penyakit Busuk Hitam (Black Rot) Daun Apel

Penyakit busuk hitam pada apel mempunyai berbagai sebutan atau penamaan yang tergantung pula pada bagian tanaman yang terserang, yakni [13]:

- Pada buah: *black rot*, *ring spot*, *brown rot*, *blossom end rot*.
- Pada daun: *frog-eye*, *brown spot*, *leaf spot*.
- Pada cabang: *cancer apel*, *cancer black rot*.

Penyakit ini pertama kali ditemukan adalah di New York, Amerika Serikat oleh Peck pada tahun 1879 sebagai busuk buah; dan pada tahun 1892 oleh Alwood sebagai bercak daun.

Gejala pada daun diawali dengan adanya bintik kecil berwarna ungu, kemudian segera meluas, dengan garis tengah bercak sekitar 2-10 mm atau rata-rata 4 mm. Bercak berbentuk bulat mempunyai garis batas yang jelas, lama kelamaan bagian tengahnya menjadi berwarna coklat kekuningan. Apabila terjadi serangan sekunder bentuk bercak menjadi tidak teratur. Pada pusat bercak dapat dilihat adanya bintik-bintik hitam yang merupakan kumpulan piknidium patogen (Gambar 3) [13].



Gambar 3. Gejala penyakit busuk hitam pada daun berupa bercak

Penyebab penyakit ini adalah jamur *Physalospora obtusa* (Schw.) Cooke, dengan sinonimnya adalah *Physalospora cydoniae* Arnaud. Patogen masuk dalam kerajaan: Fungi, filum: Ascomyca, kelas : Sardariomycetes, ordo : Xylariales, famili : Hyponectriaceae, genus : *Physalospora*. Stadium imperfectnya adalah: *Sphaeria*, *Diplodia*, *Sphaeropsis*. Akan tetapi setelah tahun 1884 Peck menyebutnya sebagai *Sphaeropsis malorum*, yang sampai saat ini nama tersebut masih sering digunakan. Travis, et.al. menyebutnya dengan nama *Botryosphaeria obtusa* (Schwein.) Shoemaker 1962 [13].

Adapun untuk parameter tingkat keparahan penyakit busuk hitam (*black rot*) pada daun apel, menurut pendapat para ahli yang disebutkan dalam penelitian oleh

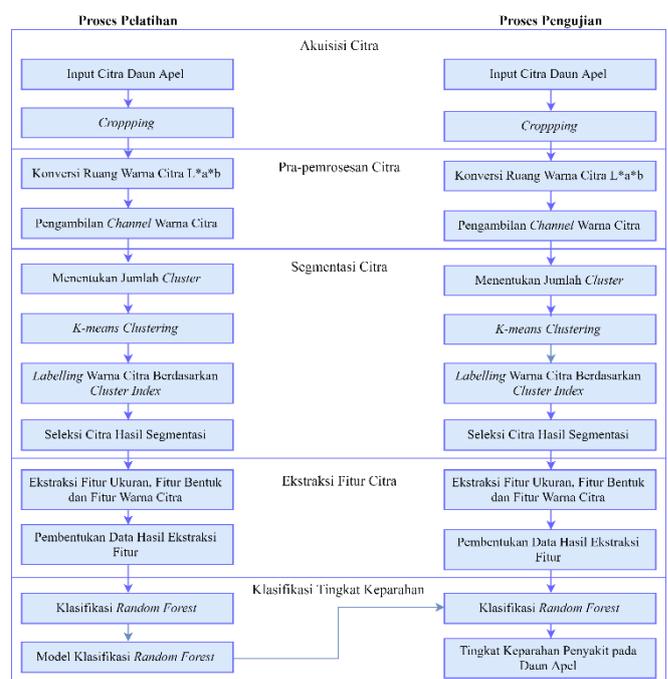
Guan Wang, Yu Sun, dan Jianxing Wang yang berjudul “Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning” pada tahun 2017 adalah tingkat keparahan di bagi menjadi empat kategori yang akan dijelaskan dalam Tabel 2 [14].

Tabel 2.
Tingkat keparahan penyakit bercak hitam (*Black Rot*) pada daun apel

Tingkat Keparahan Penyakit Bercak Hitam (<i>Black Rot</i>) pada Daun Apel	
<i>Healthy stage</i>	Tidak terdapat bagian daun yang terdapat bercak tanda terinfeksi
<i>Early stage</i>	Daun memiliki bagian yang berupa lingkaran kecil dengan diameter kurang dari 5 mm
<i>Middle stage</i>	Daun memiliki lebih dari 3 bagian yang terinfeksi dan memiliki minimal satu <i>frog-eye</i>
<i>End stage</i>	Daun memiliki bagian <i>frog-eye</i> dan bagian yang terinfeksi membesar menjadi bentuk yang tidak beraturan.

IV. PERANCANGAN DAN IMPLEMENTASI SISTEM

Perancangan sistem secara umum dijelaskan secara singkat pada Gambar 4.



Gambar 4. Diagram blok algoritma sistem

Gambaran proses secara umum dalam sistem ini terdiri dari 4 proses utama, yaitu:

- Proses pra-pengolahan merupakan serangkaian proses untuk mempersiapkan citra daun apel sebelum proses segmentasi. Proses pra-pengolahan meliputi proses *cropping* dan proses konversi ruang warna L^*a^*b .
- Proses segmentasi citra menggunakan *K-Means* merupakan tahapan untuk membedakan bagian yang terinfeksi penyakit dan bagian yang tidak terinfeksi penyakit pada citra daun apel. Pada tahap ini, citra daun apel yang telah dirubah ruang warna pada tahap sebelumnya akan dilabeli menggunakan *K-Means* dengan nilai $k=4$.
- Proses ekstraksi fitur dilakukan proses pengambilan ciri dari citra daun yang menggambarkan karakteristik citra daun apel tersebut. Fitur yang diambil dari citra

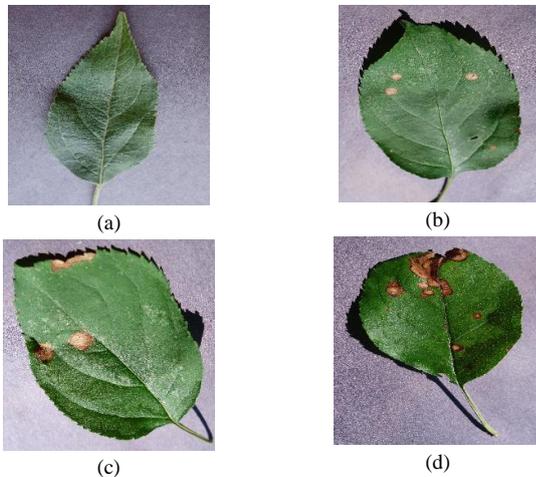
daun apel yang telah tersegmentasi adalah fitur warna, fitur luas dan fitur bentuk.

- d. Proses klasifikasi tingkat keparahan dilakukan menggunakan klasifikasi *Random Forest*.

V. UJI COBA DAN PEMBAHASAN

A. Akuisisi Data

Pada penelitian ini data input citra yang digunakan adalah citra daun apel. Contoh citra daun apel yang digunakan sebagai data masukan ditunjukkan pada Gambar 5.



Gambar 5. Citra hasil akuisisi dari empat jenis tingkat keparahan yang berbeda. (a) *Healthy Stage* (b) *Early Stage* (c) *Middle Stage* dan (d) *End Stage*.

B. Tahap Pra-pengolahan

Pada tahap ini terdapat dua proses yaitu *cropping* dan konversi warna.

a. Proses *Cropping*

Proses *cropping* yaitu proses menghilangkan *background* dan menghilangkan *shadow* (bayangan) sebelum dilakukan proses selanjutnya. Proses ini dilakukan untuk mempermudah proses-proses selanjutnya.

b. Proses Konversi Warna

Proses konversi warna yaitu proses merubah ruang warna citra dari RGB menjadi ruang warna L^*a^*b . Proses ini bertujuan untuk mempermudah pelabelan pada saat segmentasi berdasarkan warna dengan *K-Means*.

Hasil citra setelah melalui segmentasi ditunjukkan pada Gambar 6.



Gambar 6. Citra hasil pra-pengolahan. (a) Citra hasil *cropping* (b) Citra hasil konversi warna L^*a^*b .

C. Segmentasi Citra

Pada tahap ini citra daun apel yang telah melalui proses pra-prosesing, yaitu *cropping* dan konversi warna ke L^*a^*b kemudian di segmentasi menggunakan metode clustering *K-means*. Dari uji coba yang dilakukan nilai k terbaik yang digunakan adalah 4. Sehingga, nantinya diperoleh 4 citra hasil segmentasi yang selanjutnya akan diseleksi untuk

memperoleh satu citra hasil segmentasi yang diinginkan. Hasil citra setelah melalui segmentasi ditunjukkan pada Gambar 7.



Gambar 7. Citra hasil segmentasi.

D. Ekstraksi Fitur

Ekstraksi fitur menggunakan data citra hasil tahap segmentasi citra. Tahap ekstraksi fitur ini menggunakan fitur warna, luas dan bentuk yaitu *metric*, *eccentricity*, *area*, *perimeter*, *mean*, *standart deviation*, *kurtosis* dan *skewness*.

E. Klasifikasi *Random Forest*

Klasifikasi *Random Forest*, terbagi menjadi dua tahap, yaitu proses pelatihan dan proses pengujian.

1. Proses Pelatihan

Klasifikasi *Random Forest* menggunakan data hasil ekstraksi fitur. Kombinasi proporsi data juga digunakan untuk pelatihan dan pengujian, serta dilakukan pengujian klasifikasi *Random Forest* berdasarkan perbedaan jumlah pohon keputusan yang dibangun.

Pada klasifikasi *Random Forest* setiap pohon keputusan yang ditumbuhkan berdasarkan *resampling* data akan membawa *error* tersendiri pada setiap pohon keputusan. Perhitungan *error* yang diperoleh pada setiap pohon keputusan yang dibangun merupakan hasil dari *misclassified observations*. Keseluruhan *error* dari percobaan yang telah dihitung, pada tahap pelatihan klasifikasi *Random Forest* dengan kombinasi proporsi data dan kombinasi jumlah pohon keputusan, maka keseluruhan rata-rata hasil perhitungan *error* pada proses pelatihan ditampilkan pada Tabel 3.

Tabel 3. Tabel rata-rata hasil perhitungan *error*

Kombinasi Proporsi Data	Jumlah Pohon				
	10	50	100	250	500
60 : 40	0,4151	0,3443	0,3280	0,3159	0,3097
70 : 30	0,4040	0,3530	0,3383	0,3340	0,3319
75 : 25	0,4137	0,3419	0,3221	0,3045	0,2978
80 : 20	0,4279	0,3539	0,3352	0,3250	0,3198
85 : 15	0,4226	0,3468	0,3302	0,3178	0,3099
90 : 10	0,4167	0,3484	0,3314	0,3186	0,3144

Setiap model klasifikasi hasil pelatihan data memuat informasi *Variable Importance*. Keseluruhan *Variable Importance* yang dihasilkan dari semua hasil pelatihan ditampilkan pada Tabel 4.

Tabel 4. Tabel *variable importance*

Kombinasi Proporsi Data	Jumlah Pohon				
	10	50	100	250	500
60:40	Perimeter	Area	Metric	Metric	Metric
70:30	Area	Metric	Metric	Metric	Metric
75:25	Metric	Metric	Metric	Metric	Metric
80:20	Std. Dev	Metric	Metric	Metric	Metric
85:15	Metric	Metric	Metric	Metric	Metric
90:10	Metric	Metric	Metric	Metric	Metric

Berdasarkan Tabel 4. dapat dilihat bahwa mayoritas variabel yang terpilih menjadi *Variable Importance* adalah variabel *Metric*. Terkecuali pada beberapa kombinasi proporsi data dan jumlah pohon keputusan terdapat variabel *Perimeter*, *Area* dan *Standart Deviation* yang terpilih menjadi *Variable Importance*.

Pada proses pelatihan, pengukuran performa hanya didasarkan pada nilai akurasi dari model klasifikasi yang diperoleh dengan data pelatihan itu sendiri. Hasil perhitungan akurasi proses pelatihan ditampilkan pada Tabel 5.

Tabel 5.
Tabel akurasi proses pelatihan

Kombinasi Proporsi Data	Jumlah Pohon				
	10	50	100	250	500
60 : 40	98,5714	100	100	100	100
70 : 30	98,7156	100	100	100	100
75 : 25	98,5143	100	100	100	100
80 : 20	98,1818	100	100	100	100
85 : 15	98,2368	99,9496	100	100	100
90 : 10	98,7619	100	100	100	100

Berdasarkan Tabel 5 nilai akurasi tertinggi diperoleh nilai 100%. Namun, pada kolom jumlah 10 pohon keputusan yang dibangun keseluruhan proporsi data belum menunjukkan akurasi sempurna dan hanya diperoleh nilai akurasi tertinggi sebesar 99,3884% untuk proporsi data 70:30.

2. Proses Pengujian

Dari kombinasi proporsi data pengujian dan pelatihan serta perbedaan jumlah pohon keputusan seperti yang telah dijelaskan pada bagian sebelumnya, tahap pengujian proses pengujian memperoleh hasil *recall*, *precision* dan akurasi yang ditampilkan pada Tabel 6, Tabel 7 dan Tabel 8.

Tabel 6.
Tabel hasil *recall* proses pengujian

Kombinasi Proporsi Data	Jumlah Pohon				
	10	50	100	250	500
60 : 40	0,5929	0,6201	0,6227	0,6212	0,6125
70 : 30	0,5980	0,5962	0,5942	0,5874	0,5858
75 : 25	0,5985	0,6191	0,6097	0,6215	0,6144
80 : 20	0,6209	0,6471	0,6617	0,6612	0,6507
85 : 15	0,5928	0,6320	0,6111	0,6003	0,6075
90 : 10	0,6016	0,6503	0,6671	0,6502	0,6315

Tabel 7.
Tabel hasil *precision* proses pengujian

Kombinasi Proporsi Data	Jumlah Pohon				
	10	50	100	250	500
60 : 40	0,5878	0,6248	0,6393	0,6320	0,6257
70 : 30	0,5832	0,5841	0,5837	0,5752	0,5751
75 : 25	0,6119	0,6437	0,6207	0,6334	0,6365
80 : 20	0,6233	0,6640	0,6702	0,6685	0,6493
85 : 15	0,5853	0,6312	0,6273	0,6078	0,6133
90 : 10	0,6209	0,6541	0,6690	0,6613	0,6462

Tabel 8.
Tabel hasil akurasi proses pengujian

Kombinasi Proporsi Data	Jumlah Pohon				
	10	50	100	250	500
60 : 40	66,5241	69,0909	69,7326	69,5187	69,5187
70 : 30	67,8572	69,5714	69,7143	69,5714	69,4286
75 : 25	69,4017	71,2821	71,2821	72,3077	71,7949
80 : 20	69,2473	70,7527	72,0430	72,0430	71,3978
85 : 15	68	70,5714	69,1429	68,8571	69,7143
90 : 10	68,5106	73,6170	75,3191	74,8936	72,7659

Dari data Tabel 6, 7 dan 8 di atas, untuk nilai *recall* tertinggi diperoleh dari proporsi data latih dan data uji 90:10 dengan 100 pohon keputusan dengan nilai *recall* sebesar 0,6671. Selanjutnya, untuk nilai *precision* tertinggi diperoleh nilai 0,6702 pada proporsi data latih dan data uji 80:20 dengan 100 pohon keputusan. Sedangkan nilai akurasi tertinggi diperoleh dari proporsi data latih dan data uji 90:10 dengan 100 pohon keputusan serta nilai akurasi pengujian sebesar 75,3191%.

F. Pembahasan Hasil

Dari hasil yang telah diperoleh, diketahui bahwa akurasi yang diperoleh setelah proses pengujian klasifikasi menunjukkan hasil yang berbeda-beda. Performa dari klasifikasi *Random Forest* yang telah dihitung berdasarkan *recall*, *precision* dan akurasi dipengaruhi oleh beberapa hal, antara lain sebagai berikut :

1. Tingkat akurasi yang beragam dari hasil uji coba klasifikasi dipengaruhi oleh kombinasi proporsi data latih dan data uji, serta perbedaan jumlah pohon keputusan yang digunakan.
2. Pada tahap pelatihan, mayoritas nilai akurasi yang diperoleh mencapai 100%, terkecuali pada penggunaan 10 pohon keputusan yang dibangun menunjukkan nilai akurasi berkisar 98% dan pada penggunaan proporsi data latih dan data uji sebesar 85:15 dengan 50 pohon keputusan yang dibangun menunjukkan nilai akurasi sebesar 99,9496%. Sedangkan pada tahap pengujian, nilai akurasi 66-76%. Penggunaan pohon keputusan yang semakin banyak juga akan meningkatkan performa klasifikasi *Random Forest*.
3. Kombinasi proporsi data latih dan data uji serta hasil akurasi pada masing-masingnya juga menunjukkan bahwa besarnya data latih juga mempengaruhi tingkat performa klasifikasi *Random Forest*. Semakin banyak data yang dipelatihkan, maka performa model klasifikasi tersebut juga akan semakin baik karena informasi yang tersimpan pada saat proses pelatihan akan semakin banyak. Sehingga pada saat pengujian akan lebih banyak prediksi kelas yang sesuai dengan kelas kenyataannya.

VI. KESIMPULAN

Berdasarkan hasil uji coba dan pembahasan dapat diambil kesimpulan sebagai berikut:

1. Penelitian ini telah berhasil melakukan pengukuran tingkat keparahan penyakit pada daun apel dengan menerapkan *Random Forest* melalui beberapa proses, antara lain akuisisi citra, pra-pengolahan citra yang terdiri dari *cropping* dan konversi ruang warna L^*a^*b pada citra, segmentasi citra menggunakan *K-means clustering*, ekstraksi fitur warna, luas dan bentuk serta klasifikasi menggunakan *Random Forest*.
2. Kinerja metode *Random Forest* untuk mengukur tingkat keparahan penyakit pada daun apel dengan nilai akurasi tertinggi pada proses pelatihan sebesar 100% dan nilai akurasi tertinggi pada proses pengujian sebesar 75.3191%. Tingkat performa metode *Random Forest* untuk mengukur tingkat keparahan penyakit pada daun apel yaitu dipengaruhi oleh kurangnya proses pra-pengolahan citra yang mengakibatkan hasil segmentasi citra menjadi tidak sesuai, sehingga menghasilkan nilai ekstraksi fitur yang kurang tepat.

Selain itu, proporsi data yang digunakan untuk pelatihan dan pengujian, serta penggunaan jumlah pohon keputusan dalam proses pelatihan klasifikasi *Random Forest*.

DAFTAR PUSTAKA

- [1] Badan Perencanaan Pembangunan Nasional, "Apel dalam Tentang Budidaya Pertanian," 2000.
- [2] D. Putra, *Pengolahan Citra Digital*. Yogyakarta: Andi, 2010.
- [3] S. Archana and A. Sahayadhas, "Automatic Rice Leaf Disease Segmentation Using Image Processing Techniques," *Int. J. Eng. Technol.*, vol. 7, pp. 182–185, 2018.
- [4] I. Nedjar, M. E. H. Daho, and N. Settouti, "Random Forest Based Classification of Medical X-Ray Images Using a Medical Genetic Algorithm for Feature Selection," *J. Mech. Medicin Biol.*, vol. 15, no. 2, pp. 1–8, 2015.
- [5] R. Nunir, *Pengolahan Citra Digital dengan Pendekatan Algoritmik*. Bandung: Informatika, 2004.
- [6] U. Ahmad, *Pengolahan Citra Digital*. Yogyakarta: Graha Ilmu, 2005.
- [7] E. Urnaud, "Image Segmentation Using K-Means," *J. UJF UFRIMA*, pp. 1–2, 2015.
- [8] K. Gaurav and B. P. Kumar, "A Detailed Review of Feature Extraction in Image Processing System," in *Fourth International Conference on Advanced Computing & Communication Technologies*, 2014.
- [9] Y. A. Sari, R. K. Dewi, and C. Fatichah, "Seleksi Fitur Menggunakan Ekstraksi Fitur Bentuk, Warna dan Tekstur dalam Sistem Temu Kembali Citra Daun," *J. Ilm. Teknol. Informasi-ITS*, vol. 12, no. 1, pp. 1–8, 2014.
- [10] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. Morgan Kaufman Publisher, 2012.
- [11] E. Goel and E. Abhilasha, "Random Forest: A Review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 7, no. 1, pp. 251–257, 2017.
- [12] M. Dhawangkharu and E. Riksakomara, "Prediksi Intensitas Hujan Kota Surabaya dengan Matlab Menggunakan Teknik Random Forest dan CART (Study Kasus Kota Surabaya)," *J. Tek. ITS*, vol. 6, no. 1, pp. 94–99, 2017.
- [13] R. Sastrahidayat and S. Djauhari, "Penyakit Dan Hama Apel Serta Cara Pengendaliannya," Malang, 2013.