

# **Peningkatan *Random Forest* dengan menerapkan GLCM (*Gray Level Co-Occurrence Matrix*) pada Klasifikasi *Leaf Blast* Tumbuhan Padi**

**YUSUP MIFTAHUDDIN, SOFIA UMAROH, ADLEO MALIK YAMANI**

Program Studi Informatika, Institut Teknologi Nasional Bandung  
Email: [yusufm@itenas.ac.id](mailto:yusufm@itenas.ac.id)

*Received* 16 Maret 2022 | *Revised* 17 April 2022 | *Accepted* 6 Juni 2022

## **ABSTRAK**

*Penyakit leaf blast disebabkan oleh jamur yang bernama Pyricularia Grisea yang dapat menginfeksi daun padi dan menyebabkan gejala penyakit seperti bercak yang berbentuk seperti belah ketupat yang berwarna coklat yang dapat mengakibatkan kematian pada tanaman. Tingkat penyebaran penyakit leaf blast sudah meluas hingga di Indonesia yakni pada sentra-sentra produksi padi. Penelitian dilakukan untuk mengidentifikasi Daun Padi dengan ekstraksi ciri GLCM dan klasifikasinya dengan menerapkan metode Random Forest. Jumlah data uji sebanyak 200 yang terdiri dari 100 data daun padi sehat dan 100 data daun padi berpenyakit leaf blast. Penelitian menguji keberhasilan identifikasi penyakit leaf blast dan tidak berpenyakit leaf blast. Pengujian dilakukan dengan berbagai skema yaitu 40 data uji, 80 data uji, 120 data uji, 160 data uji dan 200 data uji. Pengujian menghasilkan nilai akurasi optimal pada data uji 200 sebesar 65%, recall 65%, precision 64% dan F-measure 65% dengan rata – rata pengujian waktu klasifikasi Random Forest sebesar 0.3522s.*

**Kata kunci:** *Leaf blast, Random Forest, Padi, GLCM*

## **ABSTRACT**

*Leaf blast is a disease caused by a fungus called Pyricularia Grisea which can infect rice leaves and cause disease symptoms such as brown rhombus-shaped spots that can cause plant death. The level of spread of leaf blast disease has spread to Indonesia, namely in rice production centers. The research was conducted to identify Rice Leaf with GLCM feature extraction and classification by applying the Random Forest method. The number of test data was 200 consisting of 100 data of healthy rice leaves and 100 data of rice leaves with leaf blast disease. The study tested the success of identification of leaf blast disease and not leaf blast disease. The tests were carried out with various schemes, namely 40 test data, 80 test data, 120 test data, 160 test data and 200 test data. The test resulted in the optimal accuracy value on the 200 test data of 65%, recall 65%, precision 64% and F-measure 65% with an average testing time of Random Forest classification of 0.3522s*

**Keywords:** *Leaf blast, Random Forest, Gray-level Cooccurrence Matrix, GLCM*

## 1. PENDAHULUAN

Tanaman padi (*Oryza sativa* L) adalah sumber pangan bagi penduduk dunia terutama Indonesia **(Ayub, 2020)**. Sebagai sumber pangan setengah penduduk dunia, padi harus dipelihara agar dapat menghasilkan panen beras yang maksimal. Namun karena adanya serangan penyakit pada daun padi yakni salah satunya penyakit *leaf blast* dapat membuat panen tidak maksimal.

Penyakit *leaf blast* dapat mempengaruhi pertumbuhan tanaman padi yang disebabkan oleh jamur *Pyricularia grisea* **(Ratnawati, dkk, 2020)**. Pada tahapan pembibitan dan pertumbuhan padi, *jamur ini* menginfeksi daun dan mengakibatkan gejala penyakit yang berupa berbentuk belah ketupat berwarna kecoklat coklatan yang dapat mengakibatkan kematian. **(Balai Besar Penelitian Tanaman Padi, 2015)**.

Penyakit *leaf blast* ditemukan pada tempat-tempat produksi padi **(Sudir, dkk, 2015)**. Negara-negara mengalami penurunan hasil produksi seperti Negara Jepang mengalami penurunan 20 % – 100 %, India maksimum 10 %, China sebesar 14 %, Filipina mencapai 85% dan Brasil sebesar 100%. Berdasarkan data penurunan produksi padi tersebut, produksi padi yang disebabkan oleh penyakit blast memiliki persentase yang cukup tinggi yaitu 90%. **(Wang, dkk, 2004)**

Penelitian telah banyak dilakukan dalam mengurangi pertumbuhan penyakit leaf blast salah satunya adalah mengidentifikasi awal mula penyakit sehingga tidak menyebar **(Yohannes, dkk, 2019)**. Teknologi informasi dengan menggunakan citra daun dan mendeteksi penyakit berdasarkan pada citra tersebut adalah salah satu cara mengidentifikasi awal penyakit **(Candra, dkk, 2017)**.

Dalam hal identifikasi objek berupa citra, proses ekstraksi ciri dapat mengetahui karakteristik pada citra tersebut **(Jani, dkk, 2018)**. GLCM (*Gray Level Co-occurrence Matrix*) adalah salah satu metode yang cukup efektif dalam mengekstraksi ciri pada citra yang memiliki tekstur **(Gonzalez, dkk, 2004)**. Dengan mendapatkan nilai ciri suatu citra maka deteksi citra dapat dilakukan dengan metode klasifikasi.

Algoritma Random Forest merupakan algoritma yang menerapkan proses agregasi keputusan yang berbasis klasifikasi dan regresi. **(Dhawangkara, 2017)**. Pada tahun 2001, Breiman menemukan Random Forest dan melakukan penelitian bahwa tingkat kesalahan Random Forest lebih rendah dan hasil klasifikasi yang baik dengan memiliki kehandalan dapat memproses data dalam jumlah yang sangat besar **(Breiman, 2001)**.

Penelitian ini dilakukan dengan bantuan aplikasi berbasis *desktop* dengan menggunakan *dataset* yang diperoleh dari Kaggle.com. Data yang digunakan berupa citra daun padi berpenyakit *leaf blast* dan sehat.

Ruang lingkup permasalahan penelitian ini adalah:

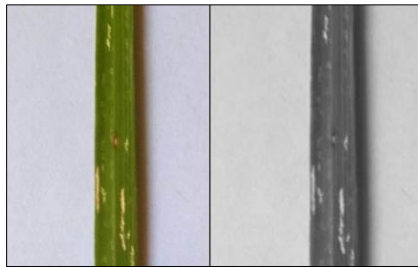
1. Klasifikasi data sebanyak 1 citra untuk setiap pengujian.
2. Klasifikasi memiliki 2 kelas yaitu padi yang berpenyakit *leaf blast* dan padi yang sehat.

Penelitian ini menerapkan ekstraksi ciri GLCM dalam mengidentifikasi *leaf blast* daun padi dan mengimplementasikan Random Forest untuk klasifikasi daun padi berdasarkan hasil identifikasi.

## 2. METODE PENELITIAN

### 2.1. Grayscale

Citra *grayscale* merupakan citra dengan range 0 – 255 yang memiliki warna abu-abu. (Indraani, dkk, 2014).



Gambar 1. Grayscale

Citra *grayscale* memiliki kanal yang memunculkan nilai intensitas (derajat keabuan). Proses *grayscale* mengubah citra yang mulanya format RGB menjadi citra grayscale dengan formula pada Persamaan (1).

$$GS(x,y) = 0.2989 * R(x,y) + 0.5870 * G(x,y) + 0.1141 * B(x,y) \quad (1)$$

Dimana :

$GS(x,y)$  = Citra *grayscale*.

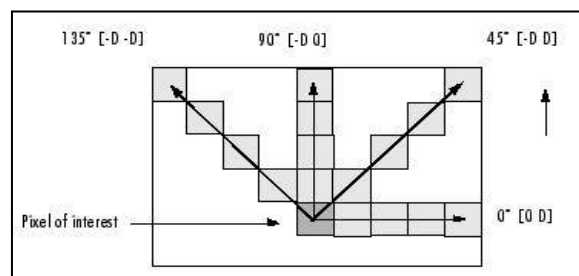
$R(x,y)$  = Nilai piksel berwarna merah (Red).

$G(x,y)$  = Nilai piksel berwarna hijau (Green).

$B(x,y)$  = Nilai piksel berwarna biru (Blue).

### 2.2 Gray Level Co-occurrence Matrix (GLCM)

GLCM berupa matriks yang menampilkan jumlah pixel  $i$  dengan pixel tetangga pada citra (Dian, 2016). Matriks kookurensi adalah matriks berukuran  $m \times m$  dimana  $m$  adalah banyaknya tingkat keabuan dengan elemen  $P(\ )$  yang *join probability distribution* pada pasangan titik-titik koordinat dengan tingkat keabuan  $x_1$  yang terletak pada koordinat  $(j,k)$  dengan  $x_2$  yang terletak pada koordinat  $(m,n)$ . Koordinat pasangan titik-titik memiliki jarak dan arah dengan sudut  $\theta$ . Arah sudut yang digunakan adalah  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , dan  $135^\circ$  yang terlihat pada Gambar 2.



Gambar 2. GLCM

Ekstraksi fitur pada GLCM menghasilkan fitur berupa nilai *Contrast*, *nilai Correlation*, *nilai Energy* dan nilai *Homogeneity* (Junita, 2017).

a. *Contrast*

*Contrast* adalah fitur yang menghitung perbedaan tingkat keabuan pada citra. Jika piksel ketetanggaan memiliki nilai yang sama, maka *Contrast* bernilai 0.

$$Con = \sum_i \sum_j (i - j)^2 p_{(i,j)} \quad (2)$$

Dimana :

Con : contrast  
i : koordinat baris pada matriks  
j : koordinat kolom pada matriks  
p : banyaknya muncul pada matriks

b. *Dissimilarity*

*Dissimilarity* merepresentasikan ukuran jarak antara pasangan objek (piksel) di wilayah yang diinginkan.

$$Diss = \sum_i \sum_j p_{(i,j)} |i-j| \quad (3)$$

Dimana :

Diss : Dissimilarity  
i : baris matriks  
j : coloum matriks  
p : kemunculan pada matriks

c. *Energy*

Fitur *energy* menghitung ukuran keberagaman citra. Semakin tinggi nilai *energy*, maka citra semakin mirip.

$$Eng = \sum_i \sum_j p_{(i,j)}^2 \quad (4)$$

Dimana :

Eng : *Energy*  
p (i,j) : kemunculan

d. *Homogeneity*

*Homogeneity* menghitung ukuran kesamaan. Jika piksel bernilai uniform maka *Homogeneity* bernilai tinggi.

$$Hom = \sum_i \sum_j \frac{p_{(i,j)}}{1+|i-j|} \quad (5)$$

Dimana :

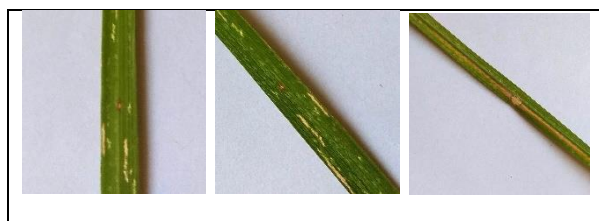
Hom : Homogeneity  
p : kemunculan  
i : baris matrix  
j : kolom matrix

### 2.3 Random Forest

*Random Forest* adalah metode klasifikasi dengan cara menggabungkan metode klasifikasi dalam hal peningkatan akurasi (Han, 2012). Metode ini menggunakan *decision tree* dengan memasukan input pada bagian atas (root) dan melanjutkan ke bagian bawah yang dinamakan daun (Muna, 2020). *Random forest* berisi kumpulan pengklasifikasi pohon yang terstruktur (Breiman, 2001). *Random Forest* memiliki sekumpulan pohon keputusan yang digunakan untuk mengklasifikasikan data. (Yohannes, dkk, 2020)

### 2.4 Leaf blast

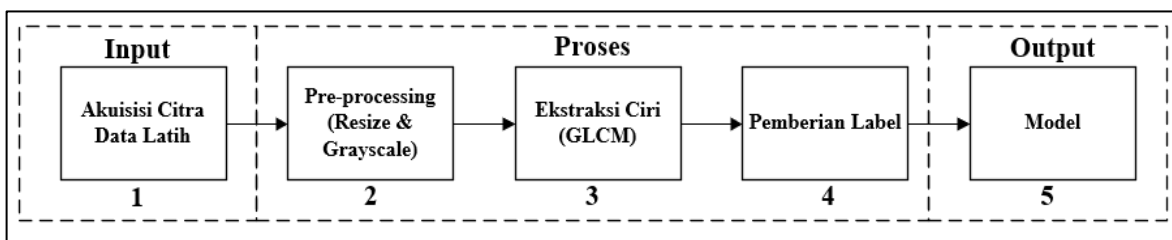
*Leaf blast* merupakan penyakit yang menginfeksi bagian daun padi dengan gejala bercak berbentuk belah ketupat berwarna coklat yang disebabkan oleh jamur *Pyricularia grisea* i. (Balai Besar Penelitian Tanaman Padi, 2015). Contoh citra daun padi berpenyakit *leaf blast* terlihat pada Gambar 3.



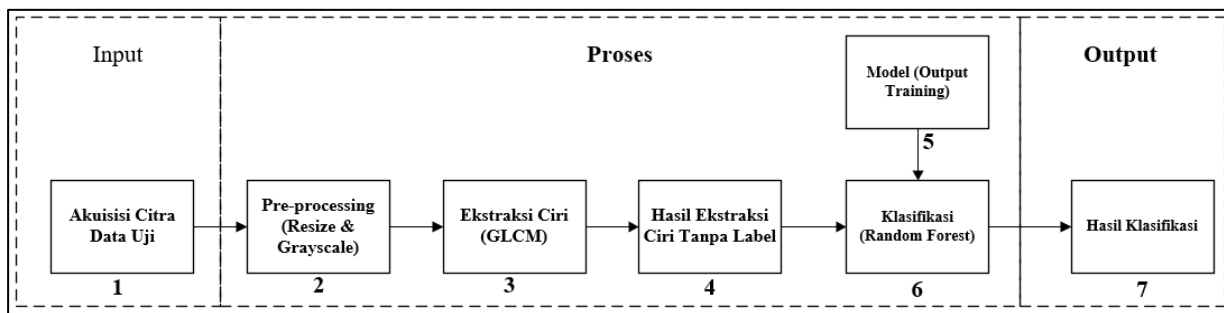
Gambar 3. Citra Daun Padi *Leaf blast*

### 2.5 Blok Diagram

Pada penelitian ini setiap dataset citra daun padi akan melalui proses sistem data latih dan uji, sistem tersebut digambarkan secara umum dengan blok diagram ditampilkan pada Gambar 4 dan Gambar 5.



Gambar 4. Arsitektur Sistem Latih

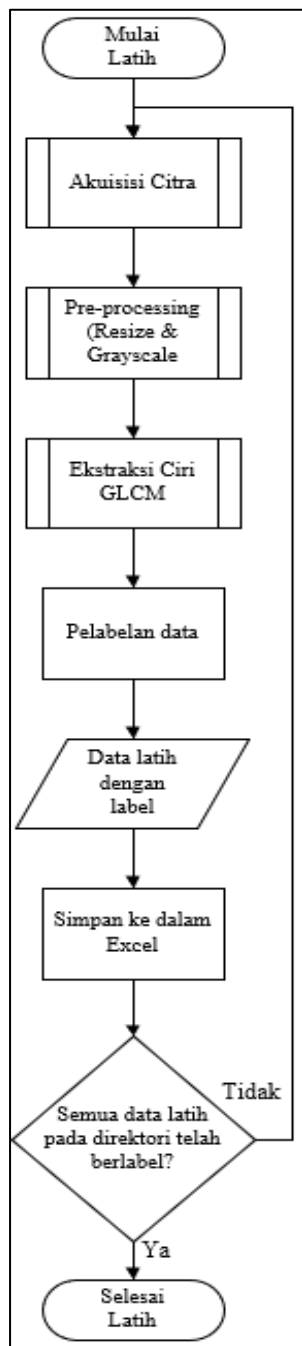


Gambar 5. Arsitektur Sistem Uji

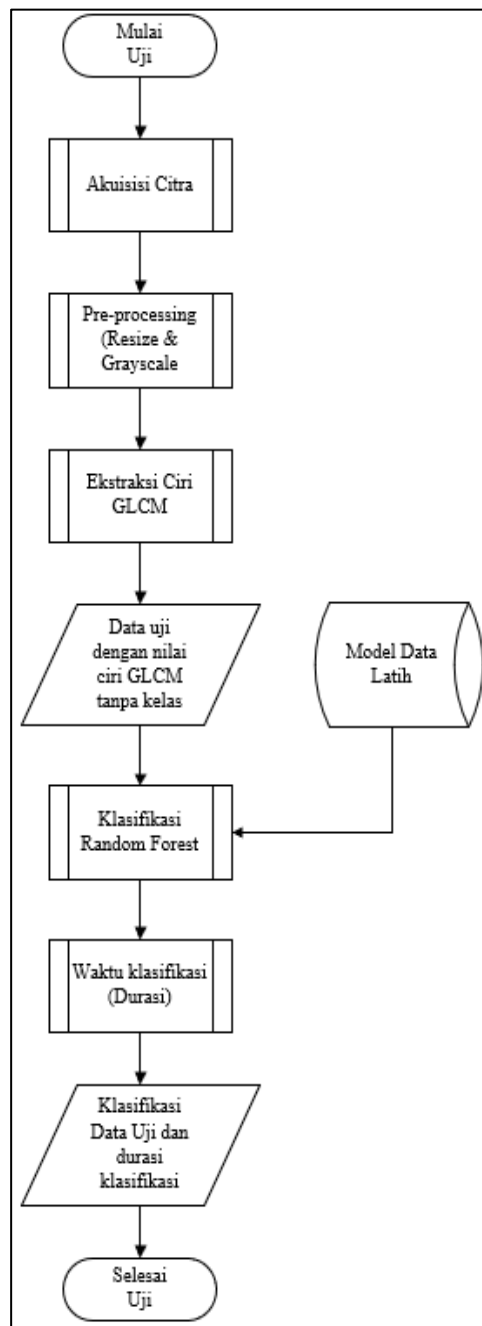
Gambar 4 dan 5 memiliki masukan citra yang kemudian citra tersebut diresize untuk mengurangi ukuran dan *grayscale* untuk mengambil nilai keabuannya. Langkah selanjutnya adalah ekstraksi ciri untuk mengambil nilai ciri setiap citra dan diberikan label.

## 2.6 Flowchart Sistem Klasifikasi Citra Padi

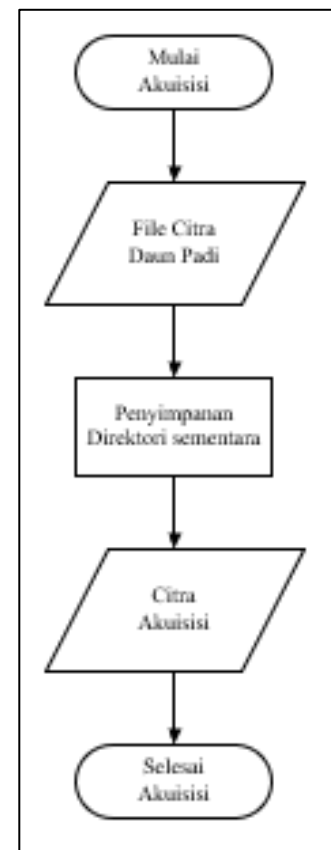
*Flowchart* ini menggambarkan alur keseluruhan sistem dimana terbagi menjadi latih dan uji untuk menentukan identifikasi dan klasifikasi *leaf blast* tumbuhan padi. Sistem dimulai dari tahap *training* data hingga proses pengujian data, pada setiap tahapan memiliki proses input citra, *preprocessing*, tahap ekstraksi ciri dan proses klasifikasi. *Preprocessing* yang digunakan untuk pengolahan citra input adalah *resize* dan *grayscale*, lalu ekstraksi ciri yang digunakan adalah ekstraksi ciri GLCM dan untuk klasifikasi yang digunakan adalah *Random Forest*. diagram alir sistem latih dan uji ditampilkan pada Gambar 6 dan Gambar 7.



Gambar 6. *Flowchart* Data Latih



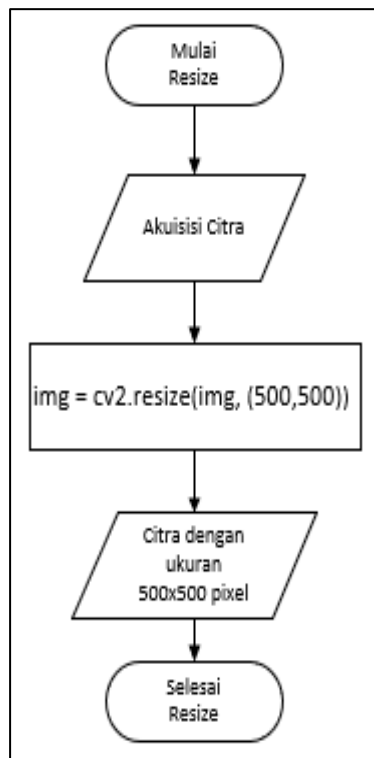
Gambar 7. *Flowchart* Data Uji



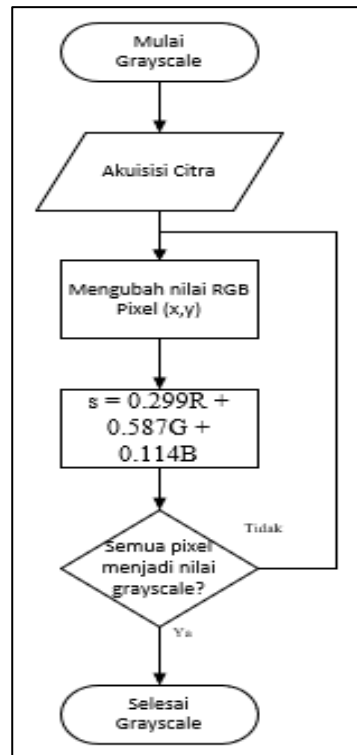
Gambar 8. *Flowchart* Akuisisi Citra

Untuk tahapan sub proses akuisisi citra digambarkan pada Gambar 8 yang merupakan diagram alir akuisisi citra, diagram alir ini merupakan sub program dari diagram alir latih dan uji.

Untuk tahapan sub proses *preprocessing* yang diilustrasikan seperti Gambar 9 dan Gambar 10 citra yang dimasukan akan dilakukan *preprocessing* dimana mengubah ukuran menjadi 500x500 piksel kemudian tahapan *grayscale* untuk mengubah nilai RGB menjadi *grayscale* sehingga nilai matriks pada citra tersebut menjadi 1 ruang bit menjadi nilai *grayscale*.



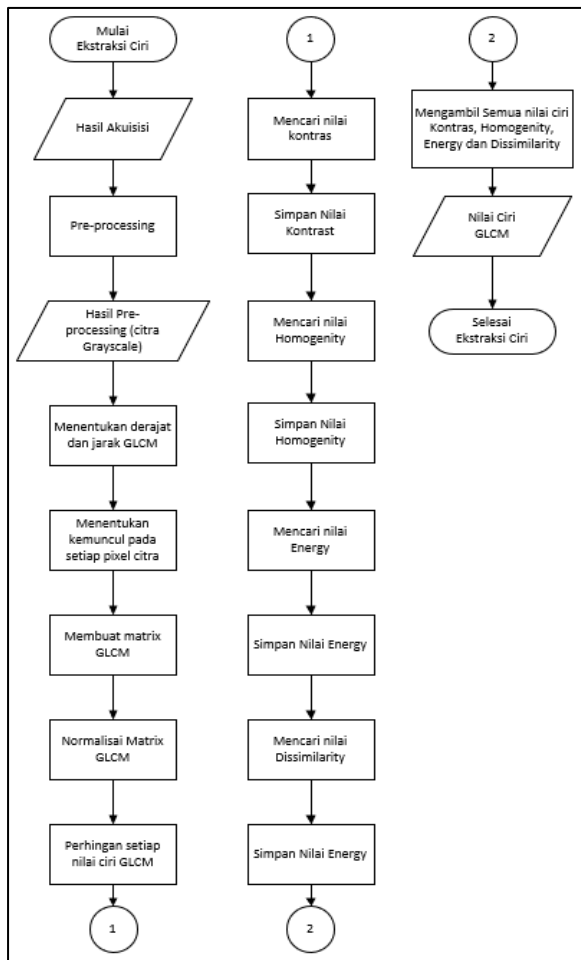
**Gambar 9. Flowchart Resize**



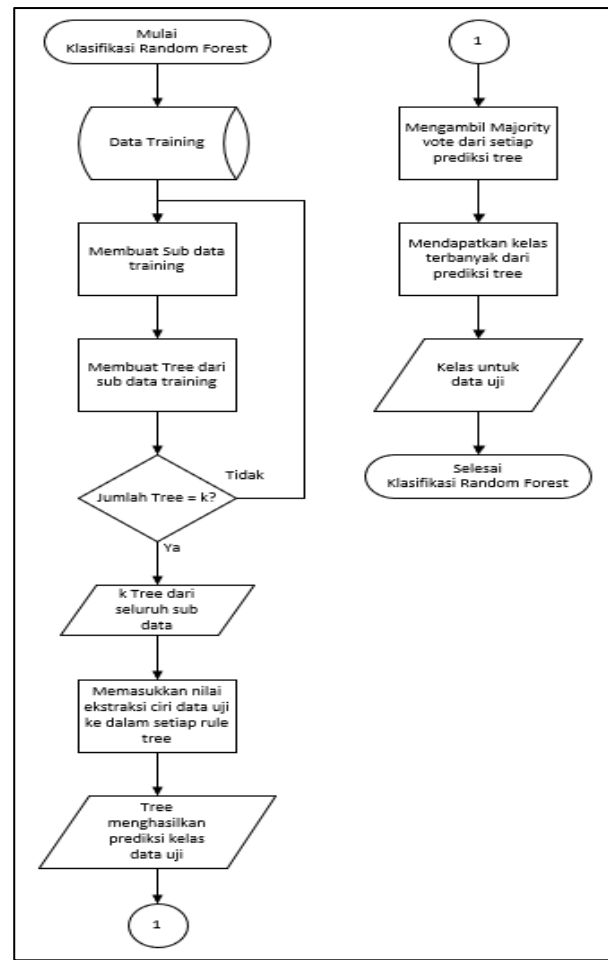
**Gambar 10. Flowchart Grayscale**

Untuk tahapan sub proses *Ekstraksi ciri* yang diilustrasikan seperti Gambar 11 sesudah citra dimasukan ke tahap *preprocessing* yang berupa tahap *resize* 500x500 piksel dan *grayscale*, citra menjadi 1 bit ruang yang dapat diproses pada tahap ekstraksi ciri GLCM.

Untuk tahapan sub proses *Klasifikasi* yang diilustrasikan pada Gambar 12 setelah citra mendapatkan nilai ekstraksi ciri, selanjutnya masuk pada tahap klasifikasi Random Forest, tahapan ini merupakan proses penentuan kelas dari citra daun padi berpenyakit *leaf blast* atau sehat (Sujaini, 2019). Prediksi kelas Random Forest menggunakan decision tree yang dibuat dari beberapa subdata dari data *training*.



Gambar 11. Flowchart GLCM



Gambar 12. Flowchart Klasifikasi RF

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. Penggunaan Dataset

Data citra pada penelitian ini sebanyak 800 data yang dijadikan data *training* dan 200 citra yang diuji yang terlihat pada Tabel 1.

Tabel 1. Data Dalam Penelitian

	Data yang digunakan	Positive Leaf blast	Negative Leaf blast
<b>Pelatihan</b>	800	400	400
<b>Pengujian</b>	200	100	100

Tabel 1 menampilkan bahwa data penelitian sebanyak 800 data latih yang terdiri dari 400 data citra positif *leaf blast* (daun yang berpenyakit *leaf blast*) dan 400 data citra negative *leaf blast* (daun yang tidak berpenyakit *leaf blast*). Untuk data uji yang digunakan sebanyak 200 data uji dengan 100 data citra *leaf blast* dan 100 data citra yang tidak memiliki *leaf blast*.





$$\textit{Precision} = \frac{TP}{TP+FP} \quad (7)$$

$$F - \textit{Measure} = 2 * \frac{\textit{Precision} * \textit{Recall}}{\textit{Precision} + \textit{Recall}} \quad (8)$$

$$\textit{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (9)$$

Keterangan :

TP = perkiraan banyaknya citra padi yang benar

TN = perkiraan banyaknya citra yang bukan citra padi yang tidak berpenyakit.

FP = perkiraan banyaknya citra yang salah.

FN = perkiraan banyaknya citra yang salah pada saat proses klasifikasi.

Pengujian kinerja sistem dilakukan pada data uji sebanyak 40, 80, 120 dan 160 dan 200 data yang diambil dari 200 data uji. Data tersebut diuji dengan dibantu aplikasi yang menerapkan random forest dengan menghasilkan *Precision*, *Recall*, *F-Measure* dan *Accuracy* sesuai dengan Persamaan 6 sampai dengan 9 yang ditampilkan pada Tabel 2 – Tabel 6.

**Tabel 2. Hasil Kinerja Klasifikasi *Random Forest* 40 Data Uji**

Random Forest					
value	Kelas	Precision	Recall	F-Measure	Accuracy
1	<i>Leaf blast</i>	50%	40%	44%	50%
0	Healty	50%	50%	55%	
Rata - Rata		50%	50%	49%	

**Tabel 3. Hasil Kinerja Klasifikasi *Random Forest* 80 Data Uji**

Random Forest					
value	Kelas	Precision	Recall	F-Measure	Accuracy
1	<i>Leaf blast</i>	56%	60%	57%	55%
0	Healty	55%	50%	53%	
Rata - Rata		55%	55%	55%	

**Tabel 4. Hasil Kinerja Klasifikasi *Random Forest* 120 Data Uji**

Random Forest					
value	Kelas	Precision	Recall	F-Measure	Accuracy
1	<i>Leaf blast</i>	65%	70%	65%	62.5%
0	Healty	61%	55%	59%	
Rata - Rata		63%	62%	62%	

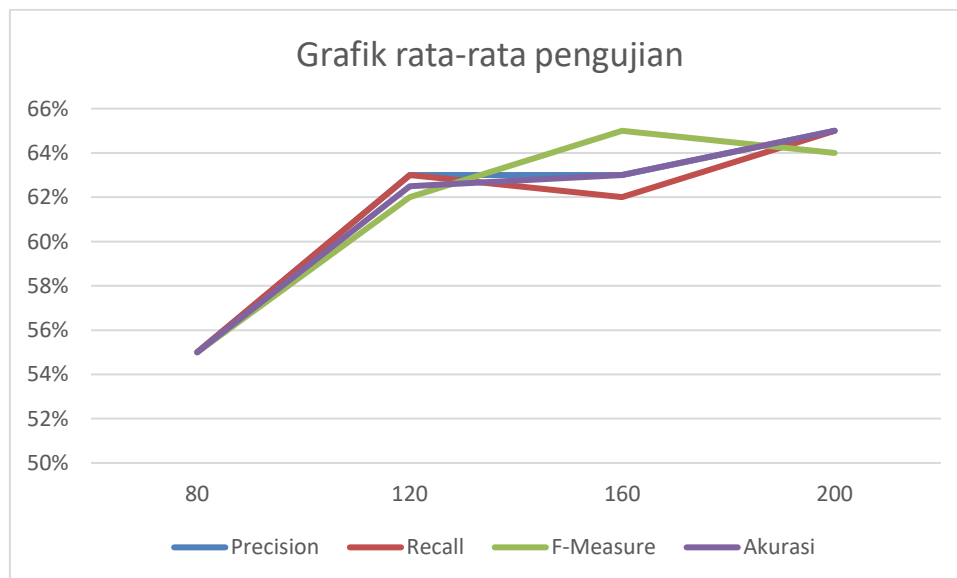
**Tabel 5. Hasil Kinerja Klasifikasi *Random Forest* 160 Data Uji**

Random Forest					
value	Kelas	Precision	Recall	F-Measure	Accuracy
1	<i>Leaf blast</i>	64%	69%	65%	63%
0	Healty	61%	56%	65%	
Rata - Rata		63%	62%	65%	

**Tabel 6. Hasil Kinerja Klasifikasi *Random Forest* 120 Data Uji**

Random Forest					
value	Kelas	Precision	Recall	F-Measure	Accuracy
1	<i>Leaf blast</i>	65%	66%	65%	65%
0	Healty	64%	63%	64%	
Rata - Rata		65%	65%	64%	

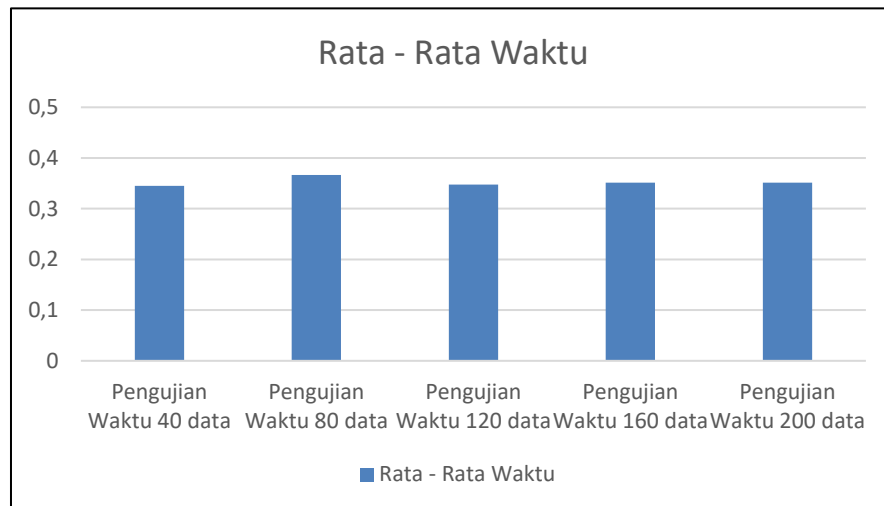
Tabel 2 sampai dengan Tabel 6 menampilkan value 1 sebagai kelas *leaf blast* dan value 0 sebagai kelas padi sehat. Akurasi diperoleh sebanyak 50 % pada jumlah data uji 40, 55 % pada jumlah data uji 80, 62.5 % pada jumlah data uji 120, 63% pada jumlah data uji 160 dan 65% pada jumlah data uji 200. Semakin banyak jumlah data uji, Random Forest dapat meningkatkan tingkat akurasi. Hal ini dikarenakan pengambilan sampel secara acak yang dibandingkan kepada dataset, sehingga semakin banyak dataset yang diuji maka semakin meningkat akurasinya.



**Gambar 14. Grafik Kinerja *Random Forest***

Gambar 14 menampilkan rata-rata pengujian dengan skema pengambilan data uji sebanyak 80, 120, 160 dan 200 dengan akurasi 65 % pada data uji sebanyak 200. Hal ini dapat dilihat bahwa nilai recall yang diperoleh yakni 65% dengan nilai persentase teridentifikasi leafblast

sebanyak 66% dan daun yang sehat sebanyak 63%, begitupun dengan nilai F-measure yang diperoleh yakni 64% dengan nilai persentase teridentifikasi leafblast sebanyak 65% dan daun yang sehat sebanyak 64%



**Gambar 15. Grafik Pengujian Waktu**

Dalam penelitian ini dilakukan pengujian waktu untuk mengukur durasi Random Forest dalam mengklasifikasi data uji didapatkan rata – rata hasil yang ditunjukkan pada grafik yang ditunjukkan pada Gambar 15. Hasil yang didapatkan adalah jumlah data tidak berpengaruh secara signifikan terhadap waktu pemrosesan.

## 5. KESIMPULAN

Pengujian dengan skema jumlah data uji sebanyak 200 data uji yang terdiri dari 100 data daun padi sehat dan 100 data daun padi berpenyakit *leaf blast*. GLCM dapat mengidentifikasi setiap data latih dan data uji yang dimasukkan dengan mendapatkan nilai cirinya. Pengujian dilakukan pada jumlah data uji sebanyak 80, 120, 160 dan 200 data uji, Random Forest dapat mengklasifikasi dan mengidentifikasi data uji ke dalam daun berpenyakit *leaf blast* atau daun padi yang sehat. Berdasarkan hasil tersebut, diperoleh nilai optimal pada data uji 200 dengan nilai akurasi 65%, *recall* 65%, *precision* 64% dan *F-measure* 65%. Pengujian juga dilakukan pada sistem dalam memproses klasifikasi dengan rata – rata pengujian waktu klasifikasi Random Forest sebesar 0.3522s. Penelitian ini menyimpulkan bahwa dapat digunakan untuk mengidentifikasi citra daun padi dan Random Forest dapat digunakan untuk mengklasifikasi citra daun padi berdasarkan hasil identifikasi GLCM.

## DAFTAR RUJUKAN

Balai Besar Penelitian Tanaman Padi. (2015, april 11). *Balai Besar Penelitian Tanaman Padi*. Diambil kembali dari <http://bbpadi.litbang.pertanian.go.id/>: <http://bbpadi.litbang.pertanian.go.id/index.php/info-berita/info-teknologi/penyakit-blas-pada-tanaman-padi-dan-cara->



- Sujaini, H. (2019). Sistem Analis Citra Alat Musik Tradisional dengan Metode k-Nearest Neighbor, Random Forest, dan Support Vector Machine. *Jurnal Sistem Informasi Bisnis*.
- Wang, X., Lee, S., Wang, J., Ma, J., Bianco, T., & Jia, Y. (t.thn.). Xi, Z. (2004). *Comparison between American and Chinese Community Building*. Dipetik May 10, 2007, dari COMM-ORG: The On-Line Conference on Community Organizing and Development: <http://comm-org.wisc.edu/papers2004/zhangxi.htm>
- Yohannes, Siska Devella, & Ade Hendri Pandrean. (2019). Penerapan Speeded-Up Robust Feature pada Random Forest Untuk Klasifikasi Motif Songket Palembang. *Jurnal Teknik Informatika dan Sistem Informasi*, Volume 5 Nomor 3.
- Yohannes, Siska Devella, & Ade Hendri Pandrean. (2020). Implementasi Random Forest Untuk Klasifikasi Motif. *Jurnal Teknik Informatika dan Sistem Informasi*, Vol. 7 No. 2.