



IDENTIFIKASI PENYAKIT TANAMAN PADA DAUN KENTANG DENGAN PENGOLAHAN CITRA MENGGUNAKAN PENDEKATAN RANDOM FOREST DAN CONVOLUTIONAL NEURAL NETWORK

by Kelompok 4

Our Team



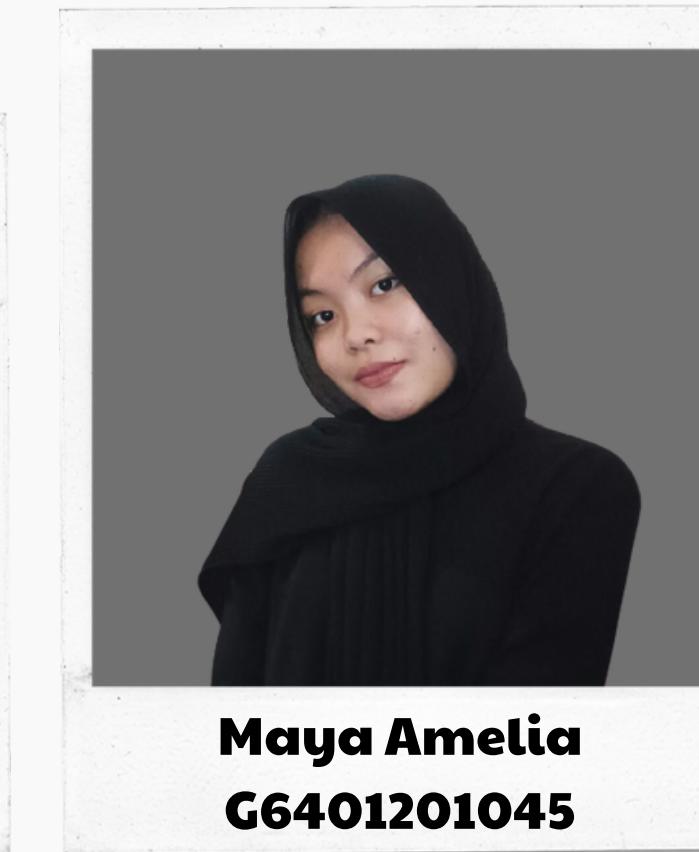
Wan Muhammad F. A
G6401201007



Helga Meilina P.F
G6401201017



Sandy Reynaldo Y. H
G6401201026



Maya Amelia
G6401201045



M. Jundi Fathan
G6401201105

LIST OF CONTENTS

01

LATAR
BELAKANG

02

RUMUSAN
MASALAH

03

TUJUAN & RUANG
LINGKUP

04

METODE

05

HASIL DAN
PEMBAHASAN

06

KESIMPULAN

LATAR BELAKANG

Kentang adalah tanaman yang banyak ditanam dan rentan terhadap berbagai penyakit yang dapat menghambat pertumbuhan dan kesehatan tanaman. Salah satu penyakit yang sering muncul pada kentang adalah hawar daun (late blight) dan early blight. Kedua penyakit ini dapat menyebabkan penurunan produksi pangan yang signifikan bagi petani. Oleh karena itu, penting untuk mengidentifikasi dan mendeteksi penyakit-penyakit ini pada tanaman kentang dengan cepat agar tindakan yang tepat dapat diambil.





RUMUSAN MASALAH

RUMUSAN MASALAH PADA PENELITIAN INI MELIPUTI:

- Bagaimana melakukan identifikasi penyakit tanaman pada daun kentang dengan menggunakan metode Random Forest dan CNN
- Bagaimana cara mendapatkan hasil identifikasi yang akurat sesuai yang diharapkan.

4

TUJUAN



Tujuan dari penelitian ini adalah melakukan analisis dan identifikasi penyakit pada tanaman daun kentang dengan menggunakan pendekatan CNN sebagai ekstraksi fitur dan Random Forest sebagai klasifikasi untuk mendapatkan hasil identifikasi yang akurat.





RUANG LINGKUP

Penelitian ini memiliki ruang lingkup dan batasan yaitu :

KELOMPOK 4

DATASET

Data diperoleh dari situs Kaggle. Di dalam data terdapat 3 buah dataset yaitu data train, valid, dan test.

ENVIRONMENT

Menggunakan bahasa pemrograman Python di google collaboration

TECHNIQUE

mengimplementasi Convolution Neural Network Algorithm untuk fitur ekstraksi dan Random Forest untuk klasifikasi

METODE



Data yang digunakan:

Potato Disease Leaf Dataset ([Kaggle](#))

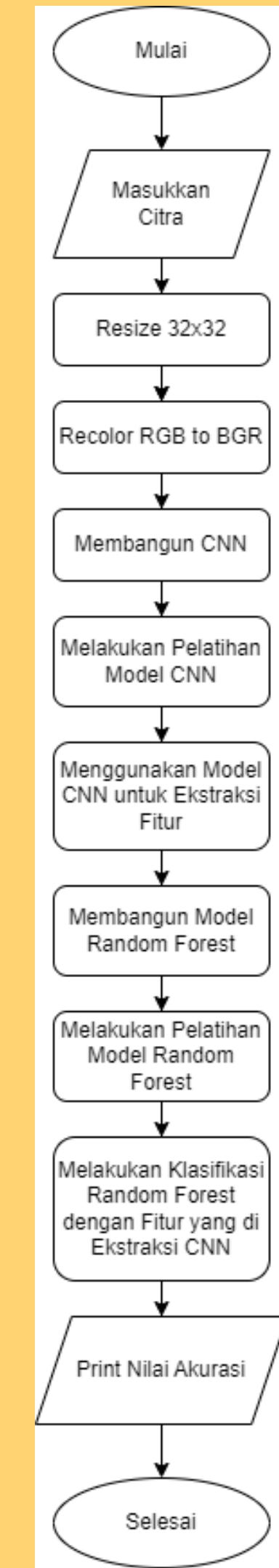
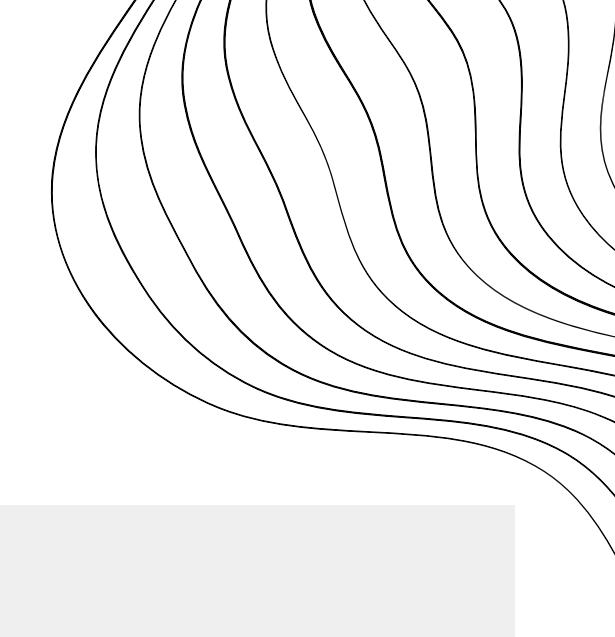
Teknik yang digunakan:

- Fitur Ekstraksi (CNN)
- Klasifikasi (Random Forest)



Lingkungan Pengembangan:

- Google Collaboratory
- Python
- Numpy, Pandas, Matplotlib, Tensorflow, Keras, Glob, Cv2



TAHAPAN PENELITIAN



PENGUMPULAN
DATA DAN PRA-
PROSES CITRA

PEMBAGIAN
CITRA LATIH
DAN CITRA UJI

FITUR EKSTRAKSI
CNN DAN
KLASIFIKASI
RANDOM FOREST

EVALUASI &
PENARIKAN
KESIMPULAN

PENGUMPULAN DATA

Data Kaggle

PLD_3_Classes_256 (3 directories)

About this directory

Potato Leaves Images Create First Time Dataset In Pakistan Region. Please Cit Research.<https://doi.org/10.3390/electronics10172064>

The screenshot shows a list of three main folders: 'Testing' (3 directories), 'Training' (3 directories), and 'Validation' (3 directories). Each folder icon contains a small folder symbol.

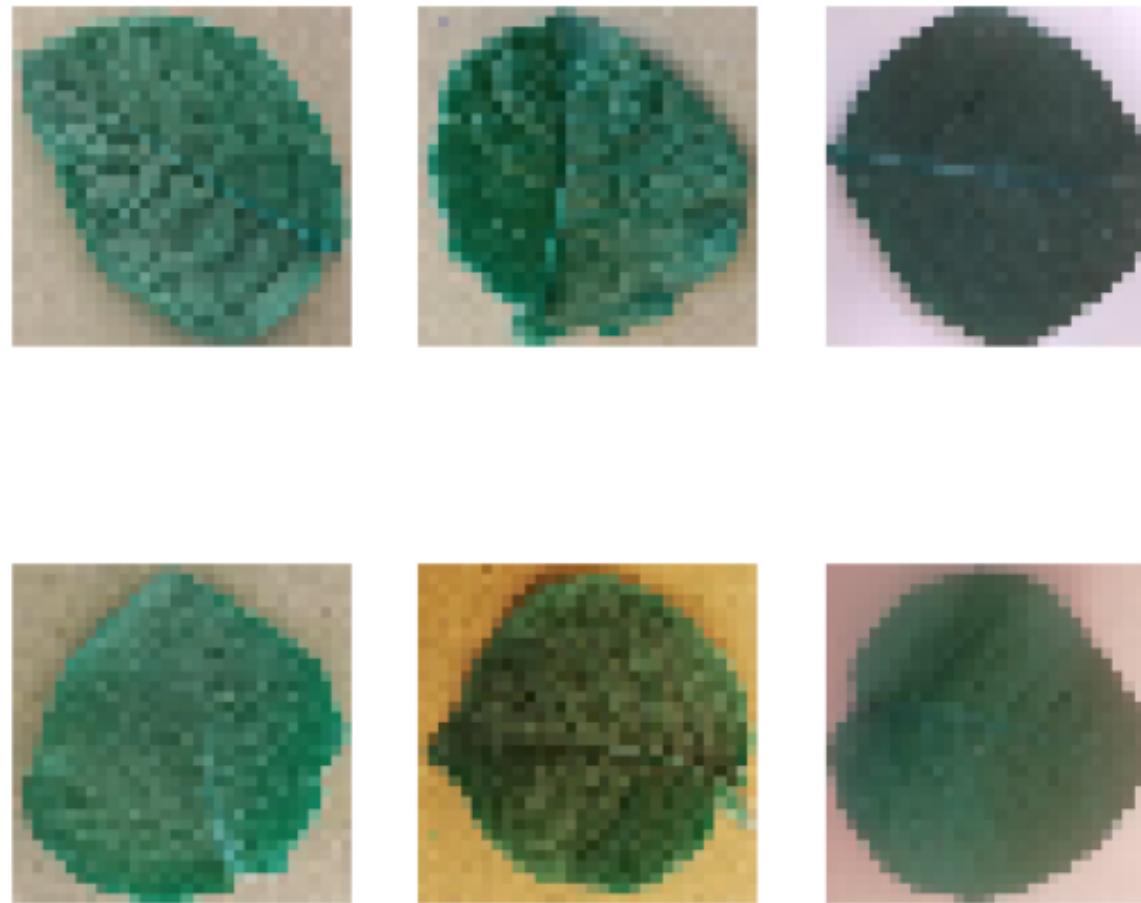
Category	Sub-Category	Count
Testing	Healthy	102
	Early Blight	162
	Late Blight	141
Training	Healthy	816
	Early Blight	1303
	Late Blight	1132
Validation	Healthy	102
	Early Blight	163
	Late Blight	151

Pengambilan data dilakukan pada kaggle yang terdiri atas 4072 gambar, terbagi menjadi 3 folder utama

- Training
 - Healthy = 102 gambar
 - Early Blight = 162 gambar
 - Late Blight = 141 gambar
- Testing
 - Healthy = 816 gambar
 - Early Blight = 1303 gambar
 - Late Blight = 1132 gambar
- Validation
 - Healthy = 102 gambar
 - Early Blight = 163 gambar
 - Late Blight = 151 gambar

PRAPROSES CITRA

Hasil Praproses



Data Training dan Data Validasi citra dilakukan praproses citra terlebih dahulu, dengan cara

- Resize citra menjadi 32x32
- Pengubahan format RGB menjadi BGR
- Melakukan labeling pada seluruh citra sesuai dengan asal foldernya

PEMBAGIAN CITRA LATIH DAN CITRA UJI

Terdapat 3251 citra yang dijadikan sebagai citra latih

```
[ ] train_leaves_images.shape, train_label_ids.shape  
((3251, 32, 32, 3), (3251,), (3251,))
```

Dan 416 citra yang dijadikan sebagai citra uji

```
[ ] test_leaves_images.shape, test_label_ids.shape  
((416, 32, 32, 3), (416,))
```

Sehingga perbandingan antara citra latih dan citra uji adalah sebesar 8:1

FITUR EKSTRAKSI CNN

Dalam melakukan ekstraksi fitur pada CNN, digunakan Model Sequential, yang menggunakan beberapa lapisan untuk memproses data. Dari lapisan Conv2D dengan 32 filter berukuran (3, 3) dan aktivasi 'relu' yang digunakan untuk mengestraksi fitur-fitur, kemudian lapisan Conv2D lain dengan 64 filter berukuran (3, 3) untuk memperdalam pemahaman fitur-fitur tersebut. Digunakan pula lapisan MaxPooling2D untuk mengurangi dimensi data tetapi tetap mempertahankan informasi penting. Kemudian, data diubah menggunakan lapisan Flatten, dan lapisan Dense sebagai lapisan output untuk melakukan klasifikasi

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 32, 32, 32)	896
conv2d_1 (Conv2D)	(None, 32, 32, 64)	18496
max_pooling2d (MaxPooling2D)	(None, 16, 16, 64)	0
flatten (Flatten)	(None, 16384)	0
dropout (Dropout)	(None, 16384)	0
dense (Dense)	(None, 256)	4194560
dense_1 (Dense)	(None, 3)	771

FITUR KLASIFIKASI RANDOM FOREST

Dalam implementasi klasifikasi Random Forest, kami menggunakan metode "fit" untuk mempelajari hubungan antara fitur-fitur dan label-label dari data pelatihan. Hal ini memungkinkan model kami untuk melakukan prediksi pada data baru. Dalam konfigurasi yang kami gunakan, kami membatasi penggunaan hingga 5 fitur dan menggunakan 50 estimator dengan random_state sebesar 40. Dengan penggunaan Random Forest, kami dapat menggabungkan kekuatan sejumlah model pohon keputusan untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting.

```
RandomForestClassifier  
RandomForestClassifier(max_features=5, n_estimators=50, random_state=40)
```

```
rf.fit(train_features, train_label_ids)
```

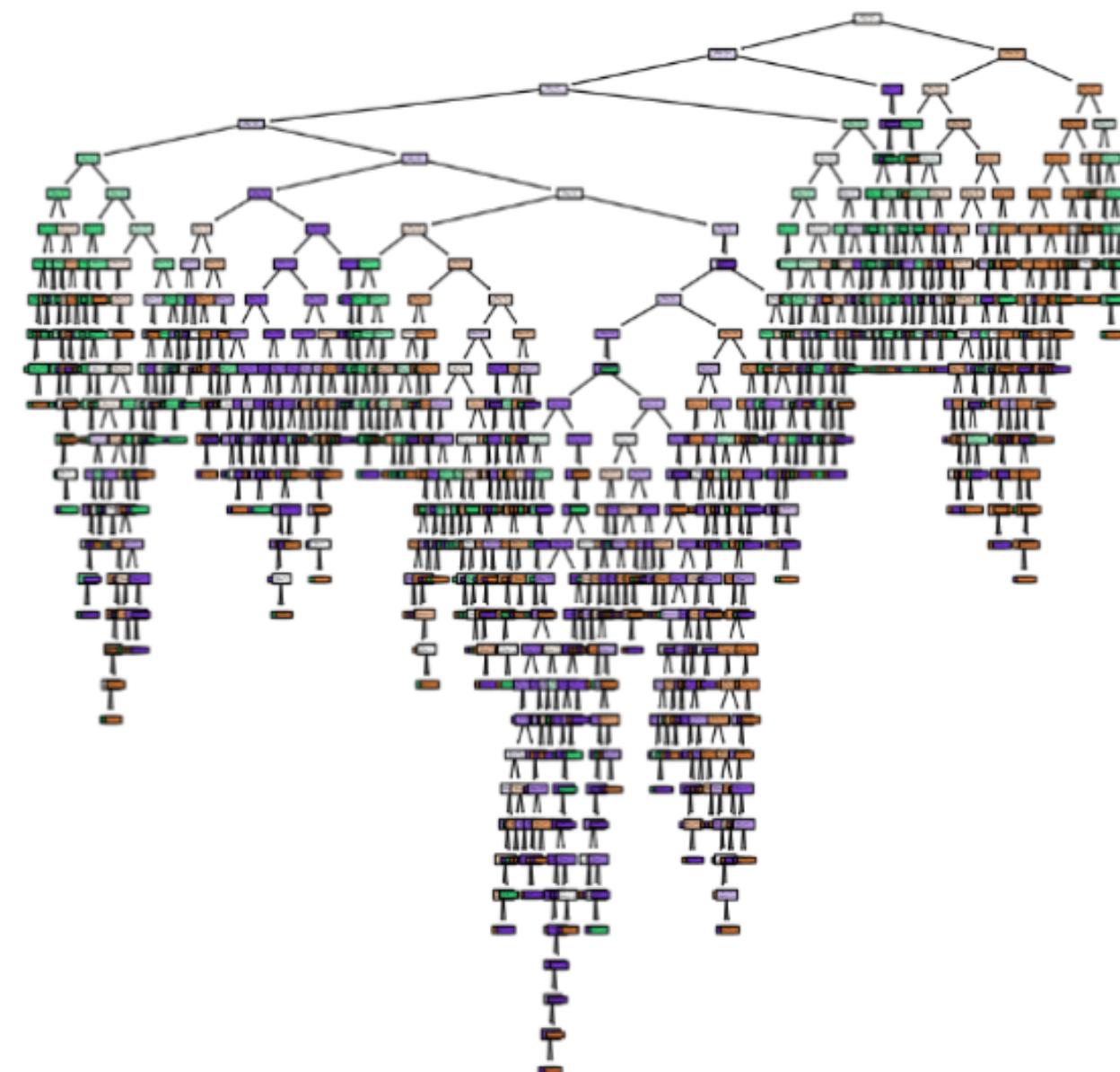
HASIL

Random forest yang dibentuk dari 256 fitur hasil ekstraksi fitur CNN

Parameter pada random forest :

- n estimator :
- random state : seed dalam random feature subsampling -> 40
- max feature : maximal fitur dalam satu decision tree -> 5

Hasil model Random Forest



ACCURACY

Akurasi random forest

```
accuracy=accuracy_score(predictions , test_label_ids)
acctrain=accuracy_score(predtrain , train_label_ids)
print('Accuracy Train:', acctrain*100, '%.')
print('Accuracy Test:', accuracy*100, '%.')

Accuracy Train: 100.0 %.
Accuracy Test: 78.125 %.
```

Akurasi

- data test : 78.125%
- data train : 100%

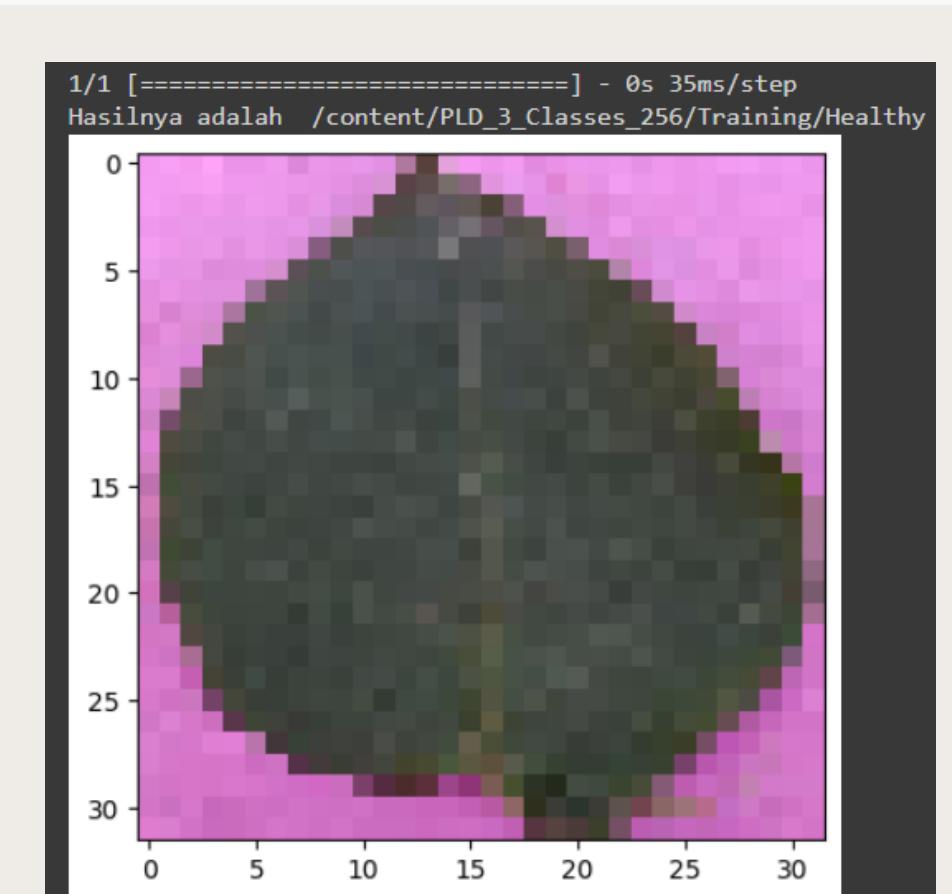
Overfitting

Solusi

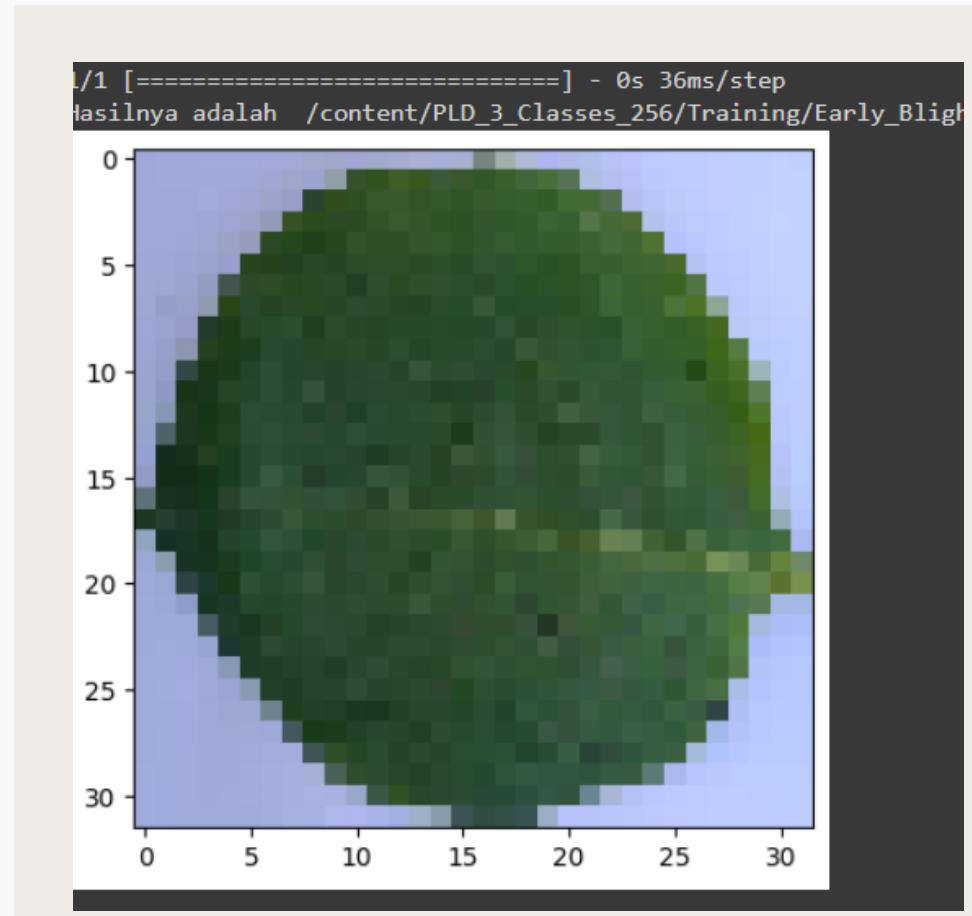
- Mencoba n-estimator, max feature, dan random state berbeda
- Meningkatkan kualitas saat melakukan pra-proses dan fitur ekstraksi (trade-off dengan komputasi)

HASIL EVALUASI

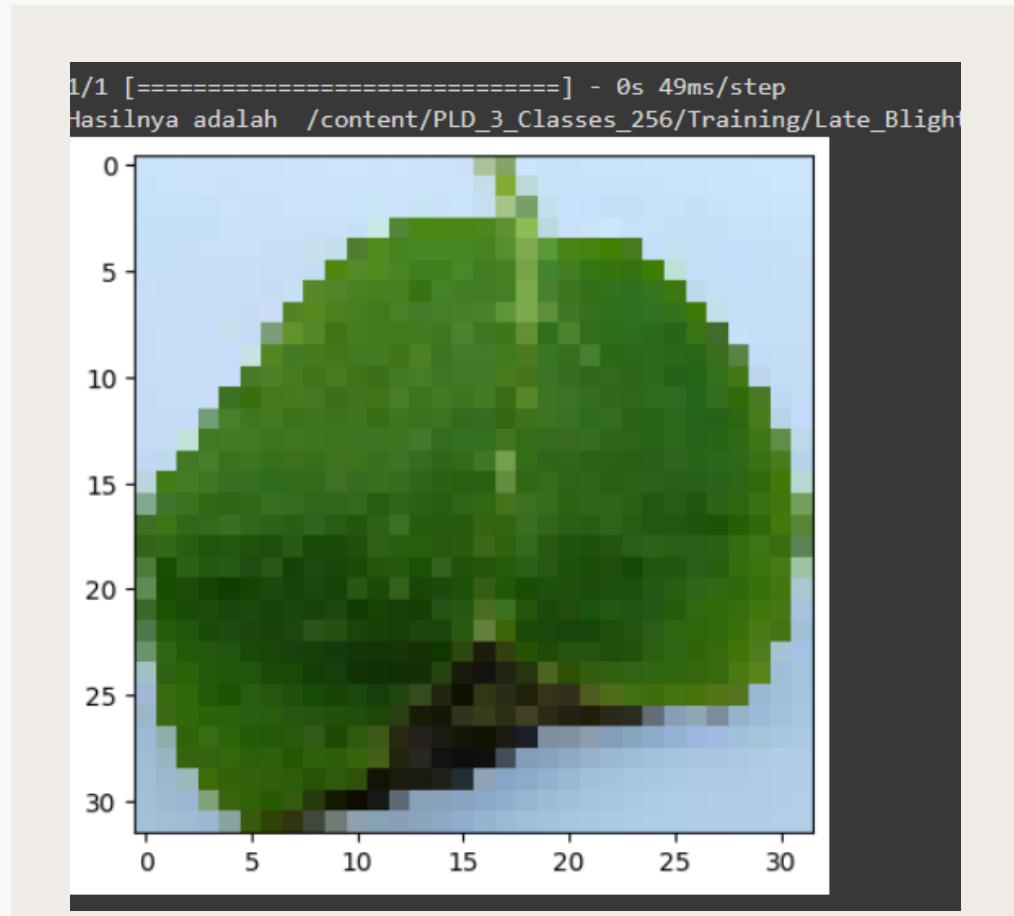
Akurasi 75% s/d 80%



Healthy



Early Blight



Late Blight



KESIMPULAN

Jadi, kesimpulan dari penelitian ini yaitu :

Pendekatan CNN memberikan kemampuan untuk mengekstraksi fitur-fitur yang relevan dari gambar daun sehingga mengenali pola dan ciri-ciri khas penyakit. Sedangkan untuk metode klasifikasi yang digunakan, yaitu Random Forest, yang dapat mengklasifikasikan penyakit dengan akurasi yang cukup baik.

Dengan analisis yang dilakukan, didapatkan hasil yang cukup akurat. Ketika dibandingkan dengan akurasi terbaik (99.75%) dan akurasi terburuk (40.05%) dari penelitian sebelumnya yang menggunakan multi-level deep learning, akurasi yang didapatkan didapatkan dengan menggunakan teknik CNN dan RF adalah sebesar 78.125%.



TERIMAKASIH

Any Question?

