

Klasifikasi Penyakit Daun Kentang Menggunakan Model Logistic Regression

Wahyutama Fitri Hidayat^{*1}, Taufik Asra², Ahmad Setiadi³

^{1,2,3}Universitas Bina Sarana Informatika

Email: ¹wahyutama.wfh@bsi.ac.id, ²taufik.tas@bsi.ac.id, ³ahmad.ams@bsi.ac.id

Abstrak

Tanaman kentang merupakan salah satu jenis umbi-umbian yang ditanam di Indonesia. Budidaya kentang dapat dikatakan tidak selalu sesuai yang diharapkan, serangan hama dan penyakit menjadi salah satu faktor penyebabnya. Sebagai upaya indentifikasi penyakit pada tanaman kentang dilakukan penelitian berdasarkan klasifikasi penyakit daun pada tumbuhan kentang. Penelitian ini berisi tentang membuat suatu sistem untuk identifikasi berdasarkan citra daun pada tanaman kentang menggunakan metode klasifikasi Logistic Regression sedangkan untuk ekstraksi fitur digunakan Resnet50. Tahap perancangan sistem diawali dengan mengumpulkan data berupa data sekunder mengenai penyakit pada daun tanaman kentang, setelah itu dilakukan fitur ekstraksi, data test dan train (pembagian data), serta menghitung nilai akurasi dan prediksi. Model ini dapat mengidentifikasi berdasarkan citra dimana sehingga menghasilkan luaran berupa nilai akurasi dari penerapan model Logistic Regression dan fitur ekstraksi Resnet50. Berdasarkan percobaan yang telah dilakukan, menggunakan data latih menghasilkan nilai akurasi sebesar 98%, sedangkan menggunakan seluruh data dengan jumlah 405 citra menghasilkan nilai akurasi sebesar 80%.

Kata kunci: *Pemrosesan Gambar, Klasifikasi, Resnet50, Logistic Regression*

Abstract

Potato plants are one type of tubers grown in Indonesia. It can be said that potato cultivation is not always as expected, one of which is caused by pests and diseases. As an effort to identify diseases in potato plants, a study was carried out based on the classification of leaf diseases in potato plants. This research is about creating a system for identification based on leaf images on potato plants using the Logistic Regression classification method, while Resnet50 is used for feature extraction. The system design stage begins with collecting data in the form of secondary data regarding diseases on potato plant leaves, after which feature extraction is carried out, data sharing (test and train data), and calculation of accuracy and prediction values. The system can later identify based on the image which will produce output in the form of predicted data results, actual data (datasets), and the accuracy value of the program. Based on the trials conducted in this study, using training data resulted in an accuracy value of 98%, while using all data with a total of 405 images resulted in an accuracy value of 80%.

Keywords: *Image Processing, Classification, Resnet50, Logistic Regression*

1. PENDAHULUAN

Kentang yang memiliki nama (*Solanum tuberosum L*) merupakan jenis umbi-umbian salah satu sumber karbohidrat selain beras, gandum, dan jagung (Hafriana & Lutfi, 2020). Membudidayakan tanaman kentang terdapat beberapa hal yang perlu diperhatikan dan tidak terlepas dari permasalahan terserang penyakit (Lesmana et al., 2022b). Penyakit utama yang dapat menyerang tanaman kentang adalah hawar daun (*late blight*) selain itu juga terdapat penyakit yang dapat menyerang pada tanaman kentang yaitu bercak kering (*early blight*) (Widestra et al., 2019).

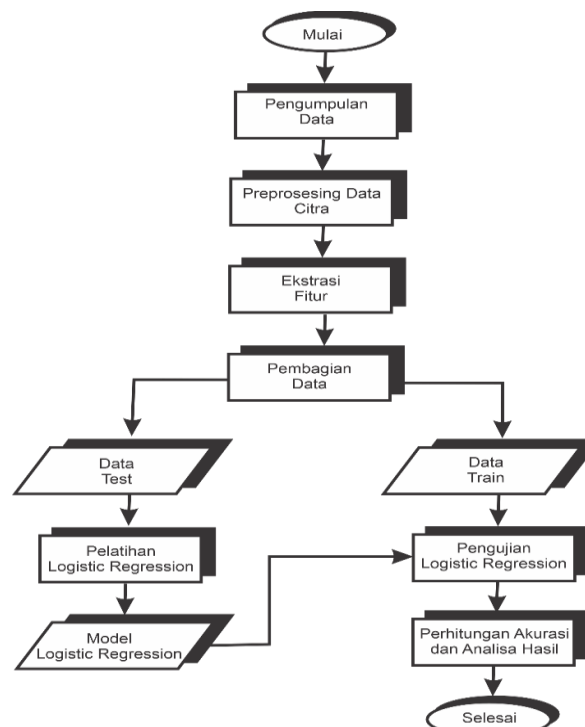
Sebagai upaya mencegah, mengidentifikasi, dan menangani penyakit tanaman kentang perlu upaya peningkatan pengetahuan untuk mengetahui tanaman yang belum atau yang sudah terserang penyakit (Lesmana et al., 2022a). Penyakit yang ada pada tanaman kentang bisa diketahui berdasarkan bercak pada daun. Bercak pada daun menjadi dasar mengidentifikasi jenis penyakit berdasarkan warna

menggunakan pengolahan citra yaitu segmentasi (Nurhidayati & Marzuki, 2020). Saat ini cara yang dalam melakukan deteksi dan identifikasi terhadap penyakit tanaman melalui pengamatan mata secara langsung (Islam et al., 2017). Namun untuk melihat gejala yang ada terdapat beberapa kendala salah satunya tidak semua petani memahami tentang cara pengenalan gejala penyakit pada daun tanaman kentang secara menyeluruh (Rozaqi et al., 2021). Penyakit yang terdapat di tanaman kentang dapat dibedakan berdasarkan dari warna dan tekstur. Berdasarkan permasalahan yang sudah diuraikan, diperlukan suatu metode yang mampu mengidentifikasi jenis penyakit pada tanaman kentang secara komputerisasi pengolahan citra daun kentang. Perkembangan teknologi pengolahan citra digital kerap kali digunakan untuk membantu permasalahan image processing (Sekar et al., 2020) dan sangat dimungkinkan memberikan efek positif terhadap kehidupan manusia dimisalkan klasifikasi pengolahan citra digital dalam pengukuran nilai akurasi (Feri Agustina, 2020). Pengolahan citra merupakan proses memanipulasi dan menganalisis citra dengan bantuan komputer (Putri, 2016). Perkembangan teknologi memiliki kemajuan yang sangat pesat dan hal itu mulai memodernisasi beberapa bidang kegiatan manusia bidang pertanian pun tak luput dari perkembangan teknologi untuk hal penelitian (Saputra & Ibadillah, 2019). Hal tersebut juga sejalan dengan penelitian pendeteksian penyakit pada tanaman (Ratnawati & Sulistyningrum, 2020).

Pengolahan citra digital harus melalui bebrapa proses agar dapat menghasilkan informasi yang diperlukan berdasarkan masukan berupa citra (Ratnawati & Sulistyningrum, 2020). Pengolahan citra merupakan teknik sebagai cara mendeteksi serta mengklasifikasi penyakit pada tanaman. Penelitian mengenai klasifikasi daun tomat pernah dilakukan oleh Lestari, dan Nauval (Lestari & Nauval, 2022) yang bertujuan untuk mengidentifikasi penyakit daun kentang menggunakan metode K-nearest neighbor (KNN) sebagai ekstraksi fitur tekstur dan warna pada tanaman kentang, penelitian ini menghasilkan nilai akurasi sebesar 80%. Penelitian lain yang dilakukan oleh Alviansyah, Ruslianto, dan Diponegoro (Alviansyah et al., 2017) klasifikasi didasarkan terhadap tiga kelas berupa daun sehat, early blight, dan late blight digunakan metode Deep Learning Convolutional Neural Network (CNN). Penelitian ini menghasilkan hasil yang baik baik pada epoch ke 10 dengan batch size 20 dihasilkan training akurasi 95% dan akurasi validasi 94%.

Berdasarkan uraian penelitian ini dilakukan identifikasi penyakit tanaman kentang dari ciri warna dan tekstur daun. Penelitian ini bertujuan untuk mengetahui nilai akurasi klasifikasi penyakit tanaman kentang menggunakan metode klasifikasi Machine Learning Logistic Regression dan fitur ekstraksi yang digunakan dalam penelitian ini dengan adalah ResNet 50.

2. METODE PENELITIAN



Gambar 1. Metodologi Penelitian

A. Pengumpulan Data

Teknik pengumpulan data atau biasa disebut dengan data collection merupakan proses dalam penelitian dimana peneliti melakukan penerapan metode ilmiah untuk mengumpulkan data secara sistematis sehingga dapat dianalisa (Karim, 2022). Tahapan ini penulis mengumpulkan dataset data latih (data train) dan data uji (data test) menggunakan data sekunder yang diperoleh dari website Kaggle.com (<https://www.kaggle.com/>) dataset "Potato-leaf".

B. Preprocessing Data Citra

Data preprocessing merupakan langkah melakukan proses awal dalam pengolahan data (Alghifari & Juardi, 2021). Pada tahap praproses data citra dilakukan dengan mengubah ukuran citra diperkecil ukurannya menjadi ukuran 224 x 224 pixel.

C. Ekstraksi Fitur

Penelitian ini menggunakan ekstraksi fitur Resnet50. ResNet-50 salah satu arsitektur bagian CNN dengan sebuah konsep shortcut connections yang berhubungan dengan keterkaitan vanishing gradient problem yang muncul saat usaha memperdalam struktur suatu network (Faiz Nashrullah et al., 2020). Menggunakan resnet50 ditransformasikan 405 gambar kedalam 405 embedding yang setiap embeddingnya memiliki 2048 features.

D. Pembagian Data

Terdapat dua tipe dalam keseluruhan data yaitu data tes (data uji) dan data train (data latih). Total keseluruhan data citra (gambar) sebanyak 405 dengan 81 citra data test , 324 citra data train.

E. Pembuatan Model Logistic Regression

Pembuatan model merupakan proses pelatihan mengenali objek dan setelah diklasifikasikan berdasarkan kelasnya (Miranda et al., 2020). Menggunakan metode klasifikasi Logistic Regression dengan studi kasus penyakit tanaman kentang akan dibuktikan bahwa akan menjadi cukup efektif dan efisien.

F. Pengujian Logistic Regression

Tahap pengujian model bertujuan untuk melatih atau melakukan proses training pada split untuk dilatih sehingga model dapat menentukan bobot untuk setiap variasi data training terhadap target (Rahman et al., 2017). Pada tahapan ini digunakan model klasifikasi Logistic Regression dan data secara keseluruhan berdasarkan 405 citra.

G. Perhitungan Akurasi dan Analisis Hasil

Penelitian ini untuk mengukur akurasi digunakan precision atau confidence dimana merupakan proporsi kasus diprediksi positif atau positif benar terhadap data yang sebenarnya selain itu juga digunakan recall atau sensitivity merupakan proporsi pada kasus positif sebenarnya diprediksi positif secara benar (Ridho & Hendra, 2022). Empat representasi hasil klasifikasi confusion matrix dimana terdapat True Positive (TP) atau data positif yang diprediksi benar. True Negative (TN) atau data negatif yang diprediksi benar. False Positive (FP) atau data negatif namun diprediksi sebagai data positif. False Negative (FN) atau data positif namun diprediksi sebagai negatif. Persamaan tabel confusion matrix sebagai berikut :



$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Penelitian ini digunakan data citra daun kentang dengan jenis daun kentang early_blight, normal, dan late_blight. Berikut merupakan contoh detail data yang digunakan disajikan pada Tabel 1.

Tabel 1. Contoh Dataset

No	Jenis Daun	Citra
1	Earliy_Blight	
2	Healthy	

3 Late_Blight



Sumber : <https://www.kaggle.com/>

B. Preprocessing Data Citra

Preprocessing data dilakukan dengan mengubah ukuran citra yang digunakan untuk penelitian. Tahap merubah ukuran dilakukan berdasarkan resolusi mendatar (horizontal) dan resolusi tegak lurus (vertikal). Tujuan merubah ukuran dimaksudkan supaya data yang digunakan dapat ditampilkan secara konsisten secara ukuran.

```
test_transforms = transforms.Compose([
    transforms.Resize([224,224]),
    transforms.ToTensor(),
])
```

Gambar 2. Proses Preprocessing Data Citra

Merubah ukuran juga dapat berdampak pada penggunaan memori menjadi lebih sedikit. Citra dirubah dimana awalnya ukuran sebenarnya 256x256 diubah menjadi 224x224.

C. Ekstraksi Fitur

Identifikasi citra penyakit pada daun kentang menggunakan kombinasi ekstraksi fitur Resnet50.

```
device = torch.device("cuda" if torch.cuda.is_available()
                      else "cpu")

feature_extractor = models.resnet50(pretrained=True)
feature_extractor.fc = nn.Sequential(nn.Linear(2048,2048))
feature_extractor.to(device)

for param in feature_extractor.parameters():
    param.requires_grad = False
```

Gambar 3. Ekstraksi Fitur

```
label_list = []
feature_list = []

labels = listdir('./dataset/')

for label in labels:
    files = listdir('./dataset/'+label+'/')
    for file in files:
        img = Image.open('./dataset/'+label+'/'+file)
        image_tensor = test_transforms(img).float()
        image_tensor = image_tensor.unsqueeze_(0)
        img_var = Variable(image_tensor)
        features_var = feature_extractor(img_var.to(device))
        features = features_var.data

    label_list.append(label)
    feature_list.append(features.cpu().numpy())
```

Gambar 4. Labeling

Model resnet50 atau biasa disebut pretrained model merupakan model yang sudah di training sebelumnya dan tinggal menggunakannya untuk menghindari error pada saat training. Hasil dari ekstraksi nantinya akan digunakan sebagai pembagian data dalam pengklasifikasian identifikasi penyakit daun kentang menggunakan metode *Logical Regression*.

D. Pembagian Data

Pembagian dibagi berdasarkan data tes (data uji) dan data train (data latihan). Pembagian digunakan perbandingan 80:20 data. Keseluruhan data citra sebanyak 405 dengan 81 citra data test , 324 citra data train.

```
X = np.array(feature_list)[: ,0,:]
y = np.array(label_list)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
print('train: ',X_train.shape, y_train.shape)
print('test: ',X_test.shape, y_test.shape)

train: (324, 2048) (324,)
test: (81, 2048) (81,)
```

Gambar 5. Pembagian Data

E. Pembuatan Model

Menggunakan metode klasifikasi Logistic Regression dengan studi kasus penyakit tanaman kentang akan diklasifikasikan berdasarkan kelasnya.

```
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression()
# pengaplikasian logistic regression
logreg.fit(X_train, y_train)
# menampilkan hasil akurasi
print('Akurasi train set dengan Logistic Regression: {:.2f}'.format(logreg.score(X_train, y_train)))
print('Akurasi test set dengan Logistic Regression: {:.2f}'.format(logreg.score(X_test, y_test)))

Akurasi train set dengan Logistic Regression: 0.98
Akurasi test set dengan Logistic Regression: 0.83
```

Gambar 6. Pembuatan Model

F. Pengujian Logistic Regression

Menggunakan metode klasifikasi Logistic Regression data yang sebelumnya sudah dibagi sehingga menghasilkan data latihan sejumlah 324 citra. Dari hasil data latihan dengan menggunakan klasifikasi Logistic Regression menghasilkan akurasi sebesar 98%.

```
thisdict = {"Model": [results_train[5][0]],
            "Accuracy": [results_train[5][1]],
            "F1 Score": [results_train[5][2]],
            "MCC": [results_train[5][3]],
            "Training Time (s)": [results_train[5][4]],
            "Predict Time (s)": [results_train[5][5]]
           }

print('Results on train data')
pd.DataFrame(thisdict)

Results on train data
```

index	Model	Accuracy
0	LogisticRegression	0.9814814814814815

Gambar 7. Akurasi Data Latihan

G. Perhitungan dan Analisa Akurasi Hasil

Menggunakan metode klasifikasi Logistic Regression data yang sebelumnya sudah dibagi sehingga menghasilkan data latihan sejumlah 324 citra dan data test sejumlah 81 citra. Dari hasil data latihan dan data test dengan menggunakan klasifikasi Logistic Regression menghasilkan akurasi sebesar 80%.

```

thisdict = {"Model": [results[5][0]],
            "Accuracy": [results[5][1]],
            "F1 Score": [results[5][2]],
            "MCC": [results[5][3]],
            "Training Time (s)": [results[5][4]],
            "Predict Time (s)": [results[5][5]]
            }

pd.DataFrame(thisdict)
    
```

index	Model	Accuracy
0	LogisticRegression	0.8024691358024691

Gambar 8. Akurasi Pengolahan Citra

4. KESIMPULAN

Setelah dilakukan penelitian, dapat disimpulkan bahwa dengan melakukan ekstraksi fitur dengan Resnet50 dan klasifikasi Logistic Regression padapenerapannya sistem mampu mengidentifikasi penyakit yang terdapat pada daun kentang. Penelitian ini digunakan data set dalam bentuk citra dengan jumlah 405 citra daun kentang. Data set dikelompokkan berdasarkan hasil pengujian ke dalam ketiga kategori yaitu Earlyy_Blight, Healthy, dan Late_Blight. Nilai akurasi yang dihasilkan sebesar 80%.

Penulis berharap kedepannya penelitian sejenis dapat lebih di kembangkan menggunakan topik pengklasifikasian pada objek lain dengan penerapan jenis ekstraksi ciri dan metode klasifikasi lainnya yang sesuai dengan kebutuhan perkembangan.

REFERENSI

- Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*, 9(02), 75–81. <https://doi.org/10.33884/jif.v9i02.3755>
- Alviansyah, F., Ruslianto, I., & Diponegoro, M. (2017). Identifikasi Penyakit Pada Tanaman Tomat Berdasarkan Warna Dan Bentuk Daun Dengan Metode Naive Bayes Classifier Berbasis Web. *Jurnal Coding Sistem Komputer Untan*, 5(1), 23–32.
- Faiz Nashrullah, Suryo Adhi Wibowo, & Gelar Budiman. (2020). The Investigation of Epoch Parameters in ResNet-50 Architecture for Pornographic Classification. *Journal of Computer, Electronic, and Telecommunication*, 1(1), 1–8. <https://doi.org/10.52435/complete.v1i1.51>
- Feri Agustina, Z. A. A. (2020). Identifikasi Citra Daging Ayam Kampung dan Broiler Menggunakan Metode GLCM dan Klasifikasi-NN. *Jurnal Infokam*, XVII(1), 25–36.
- Hafriana, D. U., & Lutfi. (2020). Analisis Kadar Glukosa Pada Kentang Rebus (*Solanum Tuberosum*) Sebagai Pengganti Nasi Bagi Penderita Diabetes Melitus Dengan Menggunakan Spektrofotometri. *Jurnal Media Laboran*, 10(1), 26–32.
- Islam, M., Dinh, A., Wahid, K., & Bhowmik, P. (2017). Detection of potato diseases using image segmentation and multiclass support vector machine. *Canadian Conference on Electrical and Computer Engineering*, 8–11. <https://doi.org/10.1109/CCECE.2017.7946594>
- Karim, R. (2022). *Teknik Pengumpulan Data, Pengertian dan Jenis*. Deepublish Store. <https://deepublishstore.com/teknik-pengumpulan-data/>
- Lesmana, A. M., Fadhillah, R. P., & Rozikin, C. (2022a). Identifikasi Penyakit pada Citra Daun Kentang Menggunakan Convolutional Neural Network (CNN). *Jurnal Sains Dan Informatika*, 8(1), 21–30. <https://doi.org/10.34128/jsi.v8i1.377>
- Lesmana, A. M., Fadhillah, R. P., & Rozikin, C. (2022b). Identifikasi Penyakit Pada Citra Daun Kentang Menggunakan Convolutional Neural Network (CNN). *Jurnal Sains Dan Informatika*, Volume 8, 21.
- Lestari, S., & Nauval, K. I. (2022). Implementasi deteksi objek penyakit daun kentang dengan metode convolutional neutral network. *Jurnal Aplikasi Teknologi Informasi Dan Manajemen (JATIM)*, 3(2), 136–149.

- Miranda, N. D., Novamizanti, L., Rizal, S., Elektro, F. T., & Telkom, U. (2020). Convolutional Neural Network Pada Klasifikasi Sidik Jari Menggunakan Resnet-50 Classification of Fingerprint Pattern Using Convolutional Neural Network in Clahe Image. *Jurnal Teknik Informatika (JUTIF)*, 1(2), 61–68.
- Nurhidayati, & Marzuki, I. (2020). Deteksi Otomatis Penyakit Daun Jagung Menggunakan Teknik Klasterisasi Data dan Operasi Morfologi. *Energy-Jurnal Ilmiah Ilmu-Ilmu Teknik*, 10(1), 1–6. <https://ejournal.upm.ac.id/index.php/energy/article/view/604%0Ahttps://ejournal.upm.ac.id/index.php/energy/article/download/604/603>
- Putri, A. R. (2016). Pengolahan Citra Dengan Menggunakan Web Cam Pada Kendaraan Bergerak Di Jalan Raya. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 1(01), 1–6. <https://doi.org/10.29100/jipi.v1i01.18>
- Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36. <https://doi.org/10.26555/jifo.v11i1.a5452>
- Ratnawati, L., & Sulistyaningrum, D. R. (2020). Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel. *Jurnal Sains Dan Seni ITS*, 8(2). <https://doi.org/10.12962/j23373520.v8i2.48517>
- Ridho, R., & Hendra, H. (2022). Klasifikasi Diagnosis Penyakit Covid-19 Menggunakan Metode Decision Tree. *JUST IT: Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 11(3), 69–75. <https://jurnal.umj.ac.id/index.php/just-it/article/view/13594>
- Rozaqi, A. J., Sunyoto, A., & Arief, M. rudyanto. (2021). Deteksi Penyakit Pada Daun Kentang Menggunakan Pengolahan Citra dengan Metode Convolutional Neural Network. *Creative Information Technology Journal*, 8(1), 22. <https://doi.org/10.24076/citec.2021v8i1.263>
- Saputra, D. E., & Ibadillah, A. F. (2019). Pengolahan Citra Digital Dalam Penentuan Panen Jamur Tiram. *Jurnal Teknik Elektro Dan Komputer TRIAC*, 6(1), 2–6. <https://doi.org/10.21107/triac.v6i1.4356>
- Sekar, P., Putri, A., Martono, K. T., & Windasari, I. P. (2020). Pengembangan Sistem Pendeteksi Gesture Angka pada Tangan secara Realtime Berbasis Android. *Edu Komputika Journal*, 7(1), 1–9.
- Widestra, S. P. G., Hidayat, N., & Dewi, R. K. (2019). Sistem Diagnosis Penyakit Tanaman Kentang Menggunakan Sistem Diagnosis Penyakit Tanaman Kentang Menggunakan Metode K-Nearest Neighbor Syndu. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(4), 4020–4026.