

SKRIPSI

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK
DETEKSI PADA PENYAKIT BAWANG MERAH
BERDASARKAN CITRA DAUN**



**Disusun untuk memenuhi sebagai persyaratan mencapai
derajat sarjana**

Disusun Oleh:

M. Apriyanto

42421041

**PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PERADABAN
2025**

SKRIPSI

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK
DETEKSI PADA PENYAKIT BAWANG MERAH
BERDASARKAN CITRA DAUN**



**Disusun untuk memenuhi sebagai persyaratan
mencapai derajat sarjana**

Disusun Oleh:
M. Apriyanto
42421041

**PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PERADABAN**

2025

PERNYATAAN PENULIS

Yang bertanda tangan dibawah ini:

Nama : M Apriyanto

NIM : 42421041

Program Studi : Informatika

Judul TA/Skripsi : Implementasi Algoritma *Random Forest* Untuk Deteksi Pada Penyakit Bawang Merah Berdasarkan Citra Daun.

Saya menyatakan bahwa laporan Tugas Akhir ini disusun berdasarkan hasil pemikiran dan pekerjaan saya sendiri. Sepanjang yang saya ketahui, karya ini belum pernah diajukan untuk memperoleh gelar akademik di perguruan tinggi manapun, serta tidak memuat kutipan atau pendapat pihak lain kecuali yang telah dicantumkan secara eksplisit dalam isi naskah dan referensinya.

Paguyangan, 09 September 2025

Yang menyatakan,



M. Apriyanto

NIM.42421041

LEMBAR PERSETUJUAN SKRIPSI

JUDUL : IMPLEMENTASI ALGORITMA *RANDOM FOREST*
UNTUK DETEKSI PADA PENYAKIT BAWANG
MERAH BERDASARKAN CITRA DAUN.
NAMA : M APRIYANTO
NIM : 42421041

Skripsi ini telah diperiksa dan disetujui oleh Dosen Pembimbing tugas akhir
guna mencapai Gelar Sarjana Komputer Strata Satu pada Fakultas Sains
Dan Teknologi Universitas Peradaban Bumiayu.

Paguyangan, 09 September 2025

Menyetujui

Pembimbing I

Pembimbing II



Tezhar Rayendra Trastanonny Pastika
Nugraha, M.Kom
NIDN. 0619019201



Khurotul Aeni, M.Kom
NIDN. 0618098802

Mengetahui

Ketua Program Studi Informatika


Nurul Mega Saraswati, M.Kom
NIDN.0606069102

LEMBAR PENGESAHAN SKRIPSI

JUDUL : IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK
DETEKSI PADA PENYAKIT BAWANG MERAH
BERDASARKAN CITRA DAUN.

NAMA : M APRIYANTO

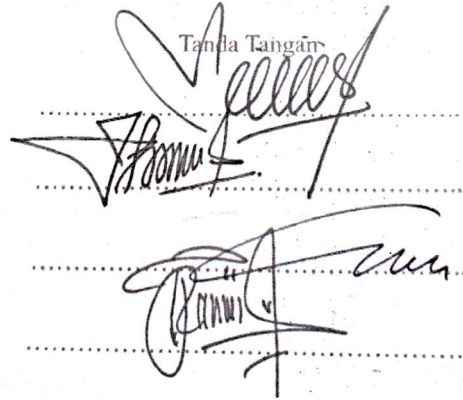
NIM : 42421041

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang
Skripsi tanggal 09 September 2025 dan dinyatakan telah memenuhi syarat

Nama Penguji

1. Nurul Mega Saraswati, M.Kom
2. Asep Saeful Millah, M.Kom
3. Khurotul Aeni, M.Kom
4. Tezhar Rayendra TPN, M.Kom

Tanda Tangan



Paguyangan 09 September 2025

Dekan Fakultas Sains dan Teknologi
Universitas Peradaban



Rizki Noor Prasetyono, M.Pd
NIDN.0611099101

Ketua Program Studi Informatika



Nurul Mega Saraswati, M.Kom
NIDN.0606069102

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul **"IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK DETEKSI PADA PENYAKIT BAWANG MERAH BERDASARKAN CITRA DAUN"**. Proposal skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika Universitas Peradaban Bumiayu.

Penulis menyadari bahwa penyusunan proposal skripsi ini tidak terlepas dari dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Dr. Muh. Kadarisman, S.H., M.Si., selaku Rektor Universitas Peradaban.
2. Bapak Rizki Noor Prasetyono, M.Pd., selaku Dekan Fakultas Sains dan Teknologi Universitas Peradaban.
3. Ibu Khurotul Aeni, M.Kom., selaku Ketua Program Studi Informatika Universitas Peradaban.
4. Bapak Tezhar Rayendra Trastaronny Pastika Nugraha, M.Kom dan Ibu Khurotul Aeni, M.Kom selaku Dosen Pembimbing yang telah bersedia memberikan waktunya untuk membimbing dan mengarahkan penulis dalam menyelesaikan laporan Skripsi ini.
5. Seluruh Dosen Informatika Universitas Peradaban Bumiayu, yang ikhlas mengajarkan ilmunya.
6. Sahabat Informatika angkatan 2021 dan seluruh angkatan yang saling mendukung dan mensupport.
7. Kedua orangtua saya, Bapak Muslih dan Ibu Taryunah yang selalu memberikan do'a dan kasih sayangnya dan selalu mendukung saya dalam keputusan yang saya pilih. Semoga beliau selalu diberi kesehatan dan panjang umur.
8. Seluruh keluarga besar saya dan kakak-kakak saya, Hopiyah, Azka, dan teman baik saya yuli terimakasih sudah memberikan suport, do'a dan

dukungannya. Semoga beliau selalu diberi kesehatan dan dilancarkan rezekinya.

9. Tidak lupa untuk semua pihak yang memberi dukungan dan suportnya kepada penulis yang tidak bisa di sebutkan satu persatu. Kami ucapkan terimakasih sebesar-besarnya.

Akhir dari skripsi penulis menyampaikan terimakasih kepada semua pihak yang telah berbagi ilmu dan pengalaman. Semoga ilmu yang sudah diberikan menjadi bekal yang bermanfaat di masa depan, semoga Allah SWT membalas segala kebaikan yang telah diberikan. Dengan selesainya skripsi ini harapan penulis semoga dapat bermanfaat bagi para pembaca, Amin

Paguyangan, 09 September 2025

Penulis



M Apriyanto

MOTTO

“Sesungguhnya bersama kesulitan ada kemudahan”

(QS. Al-Insyirah: 5-6)

ABSTRAK

Bawang merah (*Allium cepa* L.) merupakan komoditas hortikultura penting di Indonesia yang rentan terhadap penyakit daun seperti bercak ungu, moler, dan layu bakteri. Deteksi penyakit secara manual masih bergantung pengamatan visual petani yang subjektif dan kurang efisien. Penelitian ini mengimplementasikan algoritma *Random Forest* dengan ekstraksi fitur *Histogram of Oriented Gradient* (HOG) untuk mendeteksi penyakit daun bawang merah. Dataset berjumlah 165 citra dari *Kaggle* dan *Roboflow*, terbagi dalam tiga kelas, dengan pembagian data latih dan uji 80:20. Evaluasi menggunakan *confusion matrix* menghasilkan akurasi 93,94%, presisi 94,2%, recall 93,9%, dan F1-score 92,9%. Hasil tersebut menunjukkan kombinasi HOG dan *Random Forest* efektif dan stabil dalam klasifikasi citra daun serta berpotensi dikembangkan sebagai sistem deteksi penyakit otomatis untuk mendukung petani dalam pengendalian penyakit bawang merah.

Kata Kunci : Bawang merah, Citra daun, Deteksi penyakit, HOG, *Random Forest*

ABSTRACT

Shallots (Allium cepa L.) are an important horticultural commodity in Indonesia that are susceptible to leaf diseases such as purple spot, leaf spot, and bacterial wilt. Manual disease detection still relies on subjective and inefficient visual observations by farmers. This study implemented the Random Forest algorithm with Histogram of Oriented Gradient (HOG) feature extraction to detect shallot leaf diseases. The dataset consisted of 165 images from Kaggle and Roboflow, divided into three classes, with an 80:20 split between training and testing data. Evaluation using a confusion matrix yielded 93.94% accuracy, 94.2% precision, 93.9% recall, and 92.9% F1-score. These results indicate that the combination of HOG and Random Forest is effective and stable in leaf image classification and has the potential to be developed as an automated disease detection system to support farmers in shallot disease control.

Keywords: Shallot, Leaf Image, Disease detection, HOG, Random Forest

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN PENULIS	i
LEMBAR PERSETUJUAN SKRIPSI	iii
LEMBAR PENGESAHAN SKRIPSI	iv
KATA PENGANTAR	v
MOTTO	vii
ABSTRAK.....	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Penelitian Terdahulu	5
2.2 Landasan Teori	11
2.2.1 Klasifikasi.....	11
2.2.2 Machine Learning.....	11
2.2.3 Random Forest.....	12

2.2.4 Histogram of Oriented Gradient.....	14
2.2.5 Confusion Matrix.....	15
2.3 Kerangka Pemikiran	17
BAB III METODE PENELITIAN	19
3.1 Tahap Penelitian	19
3.1.1 Identifikasi Masalah	19
3.1.2 Studi Litratutur	20
3.1.3 Analisa Kebutuhan.....	20
3.1.4 Pengumpulan Data.....	20
3.1.5 Pengolahan Data.....	21
3.1.6 Training Model	22
3.1.7 Evaluasi	24
3.1.8 Hasil.....	25
BAB IV HASIL DAN PEMBAHASAN	26
4.1 Hasil Identifikasi	26
4.2 Pengumpulan data	26
4.3 Pengelolaan Data.....	27
4.3.1 Resize	28
4.3.2 Ekstraksi Fitur	28
4.3.3 Pembagian/Split Data	29
4.4 Training Model	31
4.5 Implementasi Algoritma Random Forest	32
4.6 Pengujian dan Evaluasi Model	33
4.7 Perancangan Sistem.....	37
4.7.1 Halaman Utama	37

4.7.2 Halaman Menu Dataset	37
4.7.3 Halaman Menu Modelling.....	38
4.7.4 Halaman Menu Predict.....	38
4.7.5 Halaman Menu <i>Visualize</i>	39
BAB V PENUTUP.....	40
5.1 Kesimpulan.....	40
5.2 Saran.....	40
DAFTAR PUSTAKA.....	41
LAMPIRAN – LAMPIRAN.....	43

DAFTAR GAMBAR

Gambar 2. 1 Algoritma <i>Random Forest</i>	13
Gambar 2. 2 Tahapan <i>HOG</i>	15
Gambar 2. 3 Kerangka Pemikiran	17
Gambar 3. 1 Tahapan Penelitian.....	19
Gambar 3. 2 Pengolahan Data.....	21
Gambar 3. 3 Alur Permodelan Sistem.....	23
Gambar 4.1 Distribution.....	27
Gambar 4.2 Hasil Spilt Data	30
Gambar 4.3 Hasil <i>Confusional Matrix</i>	33
Gambar 4.4 Halaman Utama	37
Gambar 4.5 Halaman Menu Dataset	37
Gambar 4.6 Halaman Menu Modeling.....	38
Gambar 4.7 Halaman Menu Predict	38
Gambar 4.8 Halaman Menu Visualize	39

DAFTAR TABEL

Table 2. 1 Tabel Perbandingan Penelitian Terkait	8
Tabel 4.1 Contoh Dataset	26
Tabel 4. 2 Hasil Extraksi Fitur.....	28
Tabel 4. 3 Pembagian Split Data	29
Tabel 4. 4 Komponen Pelatihan	31
Tabel 4. 5 Hasil Confusional Matrix	34
Tabel 4. 6 Hasil Evaluasi model.....	36

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bawang merah (*Allium cepa* L.) merupakan salah satu komoditas hortikultura penting di Indonesia yang memiliki nilai ekonomi tinggi dan peran strategis dalam memenuhi kebutuhan pangan masyarakat. Produktivitas tanaman bawang merah seringkali terhambat oleh serangan berbagai penyakit pada daun, seperti bercak ungu (*Alternaria porri*), layu bakteri (*Erwinia carotovora*), dan infeksi jamur lainnya. Serangan penyakit ini tidak hanya menurunkan hasil panen tetapi juga mempengaruhi kualitas produk yang dihasilkan. Identifikasi penyakit pada tanaman bawang merah umumnya masih dilakukan secara manual oleh petani melalui pengamatan visual terhadap gejala pada daun. Metode ini memiliki sejumlah keterbatasan, antara lain subjektivitas dalam penilaian, ketergantungan pada pengalaman individu, serta kurangnya efisiensi dalam mendeteksi penyakit secara dini. Keterlambatan dalam identifikasi penyakit dapat menyebabkan penyebaran yang lebih luas dan kerugian ekonomi yang signifikan bagi petani.

Seiring berkembangnya teknologi di bidang pengolahan citra digital dan pembelajaran mesin (*machine learning*), tersedia solusi yang lebih efisien dan akurat dalam mendeteksi penyakit tanaman. Salah satu algoritma yang efektif dalam klasifikasi citra dan data kompleks adalah *Random Forest*. Algoritma ini bekerja dengan membangun sekumpulan pohon keputusan (*decision trees*) dan menggabungkan hasilnya melalui teknik voting, sehingga menghasilkan prediksi yang akurat. *Random Forest* dalam penelitian ini didasarkan pada beberapa keunggulannya, seperti pada penelitian memiliki nilai akurasi mencapai 85% [1]. Sementara itu, penelitian menunjukkan bahwa akurasi yang dihasilkan oleh *Random Forest* lebih unggul dibandingkan *algoritma Support Vector Machine (SVM)* maupun *K-Nearest Neighbors (KNN)* dalam studi kasus klasifikasi penyakit daun padi [2]. *Random Forest* memiliki kemampuan yang baik dalam menangani data berdimensi tinggi, tahan terhadap *overfitting*, serta

tetap memberikan performa yang tinggi meskipun terdapat data yang tidak seimbang atau mengandung *noise* [3]. algoritma ini juga memberikan estimasi pentingnya setiap fitur dalam proses klasifikasi, yang sangat berguna untuk memahami kontribusi karakteristik visual daun terhadap deteksi penyakit. penelitian tentang klasifikasi penyakit bawang merah menggunakan *naïve bayes* dan *CNN* dengan fitur *GLCM*, Penelitian ini membandingkan kinerja algoritma *naïve bayes* dan *Convolutional Neural Network (CNN)* yang diektaksi menggunakan *Gray Level Co-occurrence Matrix (GLCM)* untuk mengklasifikasikan penyakit bercak ungu dan moler pada daun bawang merah dan menghasilkan akurasi 87% [4].

Penelitian klasifikasi terdahulu berkaitan dengan perbandingan algoritma *Random Forest*, *Convolutional Neural Network* dan *Support Vector Machine* untuk klasifikasi jenis mangga mendapatkan nilai akurasi tertinggi ketika menggunakan algoritma *random forest* yaitu sebesar 96%, sedangkan *Convolutional Neural Network* memperoleh nilai 83% dan *Support Vector Machine* mendapatkan nilai akurasi 95%. hasil penelitian terdahulu penggunaan algoritma *random forest* dalam deteksi penyakit pada tanaman menghasilkan akurasi 96% yang lebih tinggi. Algoritma *random forest* Sangat tepat dan bermanfaat untuk digunakan dalam studi yang disarankan untuk membantu petani mendeteksi hama pada tanaman bawang merah dengan teknologi otomatis dan presisi. Hal ini akan memungkinkan untuk mengambil tindakan lebih cepat dalam mengatasi penyakit pada tanaman bawang merah.

Penelitian dilakukan dengan menggunakan latar belakang masalah yang dijelaskan sebelumnya dapat disimpulkan bahwa *Random Forest* merupakan metode yang efektif dan efisien untuk klasifikasi penyakit tanaman berbasis citra daun, beberapa penelitian juga membandingkan dengan algoritma lain seperti *SVM* dan *Naïve bayes*, namun algoritma *Random Forest* memiliki nilai akurasi yang lebih tinggi, maka penulis melakukan penelitian dengani judul “Implementasi algoritma *Random Forest* untuk deteksi pada penyakit bawang merah berdasarkan citra daun”. Pada penelitian ini memiliki 3 klasifikasi penyakit yaitu bercak ungu, moler dan daun sehat.

1.2 Rumusan Masalah

Berdasarkan pada latar belakang dalam penelitian ini adalah: “Berapa tingkat akurasi dari model *random forest* dalam mendeteksi penyakit bawang merah?”.

1.3 Batasan Masalah

1. **Jenis Tanaman:** Penelitian ini hanya difokuskan pada tanaman bawang merah (*Allium cepa L.*) dan tidak mencakup jenis tanaman lainnya.
2. **Jenis Penyakit:** Klasifikasi penyakit daun bawang merah dibatasi pada jenis penyakit yang terdapat dalam dataset yang digunakan, tanpa menambahkan atau mengurangi jenis penyakit lainnya.
3. **Sumber Data:** Dataset yang digunakan dalam penelitian ini yaitu 165 citra daun bawang merah yang tersedia di *platform Kaggle* dan *Roboflow*, yang berisi kumpulan citra daun bawang merah dengan 3 kondisi penyakit yaitu bercak ungu, moler dan daun sehat.
4. **Metode Klasifikasi:** Proses wawancara dilakukan kepada petani bawang merah di Desa Pangebatan, Kecamatan Bantarkawung, Kabupaten Brebes. mengetahui lebih lanjut tentang penyakit apa saja yang sering dijumpai oleh petani bawang merah , ciri ciri dan cara mengatasinya.
5. **Evaluasi Model:** Evaluasi performa model klasifikasi dilakukan menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-score*. Metrik evaluasi lainnya tidak digunakan dalam penelitian ini.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi dari algoritma *Random Forest* untuk klasifikasi penyakit bawang merah.

1.5 Manfaat Penelitian

1. Bagi Peneliti

Menjadikan pemahaman yang lebih luas mengenai penerapan Algoritma *Random Forest* dalam klasifikasi penyakit bawang merah, meningkatkan analisis terhadap sebuah masalah dan bukti penerapan ilmu pengetahuan dalam bidang teknologi.

2. Bagi Universitas

Sebagai rujukan untuk penelitian lebih lanjut dalam penerapan algoritma *random forest* dalam mengidentifikasi sebuah objek.

3. Bagi Petani

Petani dapat merasakan dampak perkembangan teknologi untuk mendeteksi hama pada daun tanaman jagung dapat dilakukan secara otomatis dan akurat sehingga dapat membantu petani dalam mengidentifikasi hama pada tanaman jagung supaya dilakukan tindakan lebih lanjut dalam menangani hama.

1.6 Sistematika Penelitian

Sistematika penulisan penelitian ini dibentuk menjadi beberapa bab antara lain:

BAB I : PENDAHULUAN

Pada bab ini dituliskan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Pada bab ini berisi penjabaran teori – teori yang terkait dengan penelitian untuk memperkuat penelitian yang di buat.

BAB III : METODE PENELITIAN

Pada bab ini menjelaskan metode yang di gunakan peneliti dan tahapan tahapan yang di lakukan pada penelitian dan cara pengumpulan dataset nya.

DAFTAR PUSTAKA

Pada daftar pustaka ini berisi daftar sumber referensi yang digunakan dalam penelitian.

BAB II TINJAUAN PUSTAKA

2.1 Kajian Penelitian Terdahulu

Penggunaan metode *machine learning* untuk klasifikasi citra daun telah banyak dilakukan pada berbagai jenis tanaman, baik untuk tujuan identifikasi spesies maupun deteksi penyakit. Penelitian pertama dataset pada penelitian ini menggunakan dataset publik yang didapatkan di *kaggle.com*, memiliki 3 jenis kelas penyakit yaitu *bacterial leaf blight*, *brown spot* dan *leaf smut*. Masing-masing kelas terdiri dari 40 citra dengan ukuran 3081x897 pixel, sehingga total citra pada dataset ini adalah 120 citra. Pada penelitian ini dilakukan *augmentasi* terhadap citra asli sehingga kelas memiliki 850 citra dengan ukuran 64x64 pixel sehingga total citra yang digunakan adalah 2.550 citra. Penelitian ini membandingkan juga beberapa algoritma yaitu *Random Forest*, *KNN*, *Decision Tree*, *Naïve Bayes* dan *KNN*. Nilai akurasi yang dihasilkan oleh setiap algoritma klasifikasi diantaranya *Random Forest* dengan nilai akurasi 99.65%, *KNN* dengan nilai akurasi 99.04%, *Decision Tree* dengan nilai akurasi 98.56% dan *Naïve Bayes* dengan nilai akurasi 60.81%. Akurasi tertinggi yaitu 99.65% diperoleh ketika menggunakan algoritma *Random Forest*[2].

Penelitian kedua ini melakukan eksperimen pada data yang tidak seimbang dan data yang sudah diseimbangkan menggunakan klasifikasi *Random Forest*, *KNN* dan *Gaussian NB*. Data yang digunakan berjumlah 663.228. Data tersebut merupakan data tidak seimbang. Maka dari itu data akan diseimbangkan menggunakan algoritma *Tomek Link Undersampling*. Dari data yang tidak seimbang dan data yang telah diseimbangkan akan dilakukan pemodelan menggunakan *K-Nearest Neighbor (KNN)*, *Gaussian Naïve Bayes* dan *Random Forest*. Dari data tersebut akan dibagi menjadi data training dan data testing dengan perbandingan 80:20. Algoritma yang paling tepat untuk menyelesaikan permasalahan *Imbalanced Data* maupun *Balanced Data* menggunakan *Tomek Links Undersampling* yaitu algoritma *Random Forest*. Hasil eksperimen menunjukkan hasil performa terbaik pada algoritma *random forest* dengan nilai

akurasi tertinggi yaitu 0.9901, Algoritma *KNN* dengan nilai akurasi 0.9851, dan *Gaussian NB* 0.9852. [3].

Penelitian ketiga memiliki tujuan membandingkan kedua algoritma *CNN* dan *Naive bayes* dengan *improvement* ekstraksi fitur *FD*. Perbandingan dari kedua algoritma akan digunakan untuk membantu mengklasifikasikan citra dari 2 variabel bawang merah. Variabel penelitian ini yaitu moler dan bercak ungu. penelitian ini memiliki 320 citra dari dua variabel yaitu moler dan bercak ungu masing-masing variabel memiliki 160 citra berformat JPG dengan ukuran 1:1. Penelitian ini menghasilkan nilai akurasi terbesar ketika menggunakan algoritma *CNN* dan Ekstraksi fitur *FD* yaitu sebesar 98%, sedangkan menggunakan *naive bayes* memperoleh nilai 95%. [4].

Penelitian keempat ini membandingkan tiga metode, yaitu *random forest*, *Support Vector Machine (SVM)*, dan *Connvolutional Neural Network (CNN)* untuk mendapatkan hasil terbaik pada klasifikasi jenis mangga berdasarkan citra. Dataset yang dilakukan dalam penelitian ini berjumlah 1.600 citra yang terdiri dari 8 variabel yaitu *anwar ratool*, *chaunsa – black*, *chaunsa – summer bahist*, *chausa – white*, *dosehri*, *fajri*, *langra*, dan *sindhri*. data peneliti dilakukan pra-pemrosesan dimana citra gambar dilakukan penyeragaman ukuran citra menjadi 300x300 pixel dan dilakukan perubahan warna menjadi skala abu-abu (*grayscale*). Kemudian dibagi data latih dan data uji dengan rasio 70:30. Hasil dari perbandingan metode *random forest* sangat baik digunakan dalam pengklasifikasi jenis mangga berdasarkan citra dengan nilai akurasi 96% sedangkan pada algoritma *SVM* memiliki nilai akurasi 95% dan metode *CNN* mendapatkan nilai akurasi 83% [5].

Penelitian Kelima bertujuan untuk mengolah dan melakukan analisa data, bagaimana penerapan metode *Decision Tree*, *Naive Bayes* dan *Random Forest* pada klasifikasi penyakit daun mangrove, dalam penelitian ini terdiri dari dua jenis kelas, yaitu daun sakit dan daun sehat. masing-masing kelas terdiri dari 80 citra sehingga total citra pada dataset ini adalah 160 citra dengan ukuran 3024x4032 pixel. Pada awalnya, setiap kelas memiliki 80 gambar, dengan menggunakan teknik augmentasi ada peningkatan dengan setiap kelas memiliki

480 gambar. proses peningkatan ini agar pelatihan pengujian model dapat mempelajari lebih banyak gambar untuk menghasilkan hasil yang lebih akurat. Kemudian penelitian ini dapat disimpulkan bahwa terdapat 4 kondisi berbeda dalam memprediksi model yaitu kondisi *true positives*, *false positives*, *true negatives* dan *false negatives*. Hasil evaluasi performa model *Random Forest* pada penelitian ini yaitu 85% akurasi, 88% *specificity*, 83% *Recall*, 87% *pressi*, serta 84% nilai *F1-Score* [1].

Kesimpulan dari penelitian sebelumnya, *random Forest* secara konsisten memberikan performa klasifikasi yang tinggi pada berbagai domain, baik pengolahan citra maupun data tabel. Pada kasus klasifikasi citra daun (penelitian pertama dan keempat), *Random Forest* mampu mencapai akurasi hingga 99,65% dan 96%, metode *Random Forest* banyak digunakan untuk klasifikasi pada citra daun karena kemampuannya yang baik dan mengklasifikasi sebuah objek citra digital, serta ketahanan yang kuat dalam mengatasi data pada sebuah objek citra digital membuat algoritma *random forest* sangat efektif dalam memproses citra digital. Penelitian yang akan di gunakan selanjutnya menunjukan perbedaan dari penelitian yang dilakukan sebelumnya. Data penelitian ini menggunakan 165 daun bawang merah dengan 3 kelas yaitu moler, bercak ungu dan daun sehat. Penelitian ini menggunakan ekstraksi fitur menggunakan *HOG* kemudian memprediksi penyakit menggunakan model *random forest* untuk mendeteksi penyakit pada daun bawang merah yang diharapkan dapat membantu petani dalam melakukan deteksi penyakit sehingga langkah pengendalian dapat dilakukan otomatis dan akurat untuk dilakukan tindakan lebih lanjut. Perbedaan penelitian terkait dapat di lihat pada Tabel 2.1.

Table 2. 1 Tabel Perbandingan Peneltian Terkait

No	Penulis	Tujuan	Metode	Hasil	Perbandingan
1	Sarifah Agustiani dkk.(2022)	Klasifikasi Penyakit Daun Padi	<i>Random Forest dengan Ekstraksi fitur Color Histogram</i>	Hasil akurasi yang diperoleh dari penelitian ini adalah 99. 65%.	Penelitian sebelumnya : • Objek Daun Padi • 120 Citra • Terbagi 3 kelas • Ekstraksi Fitur Color Histogram penelitian selanjutnya : • Objek daun bawang merah • 165 dataset • Terbagi 3 Kelas • Random forest dan Extraksi fitur <i>HOG</i>
2	Ni Made Dina Aprilianti dkk. (2023)	Mencari akurasi terbaik dari ketiga algoritma tersebut dengan kondisi data yang seimbang maupun tidak seimbang	<i>KNN, Gaussian Naïve Bayes, Random Forest dan Tomek Link Undersampling</i>	Nilai akurasi tertinggi pada data tidak seimbang didapatkan dari algoritma <i>Random Forest</i> yaitu sebesar 99.6% sedangkan pada data seimbang juga random forest masih jadi tertinggi yaitu 99.9%	Penelitian sebelumnya : • Dataset Numerik (Hasil Pengukuran kimia dari ekstrak tanaman keladi Tikus) • Data yang digunakan 663.228 • Terbagi 2 kelas penelitian selanjutnya : • Objek daun bawang merah • 165 dataset • Terbagi 3 Kelas

No	Penulis	Tujuan	Metode	Hasil	Perbandingan
					Random forest dan Extraksi fitur <i>HOG</i>
3	Dian Azurah Dkk (2023)	Mengklasifikasikan penyakit bawang merah dengan mengimplementasikan metode <i>naïve bayes</i> (<i>gaussian, bernoulli, dan multinomial</i>) dan CNN pada citra bawang merah yang di ekstraksi menggunakan <i>Fourier descriptor</i> .	<i>CNN</i> dan <i>Naïve Bayes</i>	Metode <i>FD – CNN</i> memperoleh akurasi terbesar yaitu 98%, sedangkan tanpa ekstraksi <i>FD</i> menghasilkan nilai 97%, sedangkan menggunakan algoritma <i>Naïve bayes</i> menghasilkan 95%.	Penelitian sebelumnya : - Objek daun bawang 320 citra daun bawang <i>CNN</i> dan <i>Naïve bayes</i> - Terbagi 2 kelas penelitian selanjutnya : - Objek daun bawang merah 165 dataset - Terbagi 3 Kelas - Random forest dan Extraksi fitur <i>HOG</i>

No	Penulis	Tujuan	Metode	Hasil	Perbandingan
4	Ricky Mardianto dkk. (2024)	Klasifikasi Jenis Mangga dengan mencari nilai akurasi terbaik dari ketiga algoritma tersebut	<i>Random Forest, CNN, Support Vector Machine (SVM).</i>	Penggunaan random forest pada penelitian ini mendapatkan nilai akurasi tertinggi yaitu 96%, sedangkan <i>precision</i> , <i>recall</i> dan <i>f1score</i> sebesar 96.5%.	Penelitian sebelumnya : • Objek Jenis Mangga • 200 Citra Jenis Mangga • <i>Random Forest, CNN, Support Vector Machine (SVM).</i> • Terbagi 8 kelas penelitian selanjutnya : • Objek daun bawang merah • 165 dataset • Terbagi 3 Kelas • Random forest dan Extraksi fitur <i>HOG</i>
5	Rizky Muhammad (2024)	Klasifikasi kesehatan daun mangrove	<i>Random Forest</i>	Menghasilkan nilai akurasi sebesar 85%	Penelitian sebelumnya : • Objek aun bawang • Dataset 320 • Terbagi 2 kelas penelitian selanjutnya : • Objek daun bawang merah • 165 dataset • Terbagi 3 Kelas • Random forest dan Extraksi fitur <i>HOG</i>

2.2 Landasan Teori

2.2.1 Klasifikasi

Klasifikasi adalah proses pengelompokan objek ke dalam kelas-kelas berdasarkan ciri atau fitur yang dimiliki. Dalam konteks pengolahan citra digital, klasifikasi bertujuan untuk mengidentifikasi jenis objek yang terekam dalam gambar. Pada penelitian ini, objek yang diklasifikasi adalah citra daun bawang merah yang mengalami gejala penyakit. Untuk itu, diperlukan metode klasifikasi yang andal dan akurat dalam mengenali pola visual yang kompleks. Metode klasifikasi bekerja dengan membangun model berdasarkan data latih, kemudian menggunakan model tersebut untuk mengklasifikasikan data baru. [1]. Klasifikasi termasuk ke dalam kategori *supervised learning*[6]. Dalam proses klasifikasi, sebuah model dibangun dari data latih (*training data*) yang berisi sekumpulan data input beserta label kelas yang benar. Model yang telah dilatih ini kemudian digunakan untuk memprediksi label dari data baru yang belum diketahui kelasnya. Proses ini sangat berguna dalam berbagai bidang, seperti deteksi penyakit, dan klasifikasi teks atau citra.

2.2.2 Machine Learning

Suatu kumpulan teknik yang memungkinkan penggunaanya untuk melakukan implementasi algoritma dimana memiliki sifat dapat beradaptasi untuk mengorganisir data masukkan berdasarkan fitur yang sama secara otomatis dan juga melakukan prediksi dikenal dengan nama Machine Learning [1]. *Machine Learning* (pembelajaran mesin) juga termasuk cabang dari kecerdasan buatan (*Artificial Intelligence/AI*) yang memungkinkan komputer untuk belajar dari data dan membuat keputusan atau prediksi tanpa diprogram secara eksplisit [7]. Dalam konteks ini, "belajar" berarti sistem secara otomatis meningkatkan performanya berdasarkan pengalaman (data historis).

a. *Supervised Learning*

Supervised learning adalah metode pembelajaran di mana algoritma dilatih menggunakan data yang telah diberi label. Tujuan utamanya

adalah membangun model yang dapat memetakan input ke output yang benar[8].

Contoh: klasifikasi penyakit tanaman, deteksi spam email.

Contoh Algoritma :

- 1) *Decision Tree*
- 2) *Random Forest*
- 3) *Support Vector Machine (SVM)*
- 4) *Naïve Bayes*
- 5) *K-Nearest Neighbor (KNN)*

b. *Unsupervised Learning*

Unsupervised learning digunakan untuk menemukan struktur tersembunyi dalam data yang tidak memiliki label. Pendekatan ini umum digunakan untuk pengelompokan data (*clustering*) dan reduksi dimensi[8].

Contoh Algoritma :

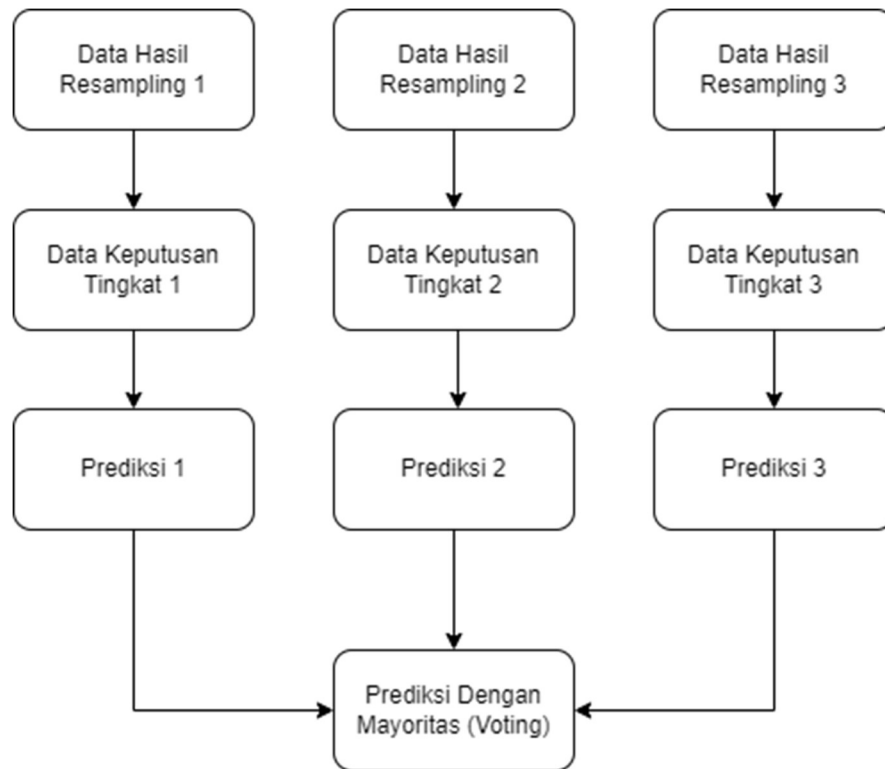
- 1) *K-Means Clustering*
- 2) *Principal Component Analysis (PCA)*
- 3) *DBSCAN*

2.2.3 Random Forest

Random Forest adalah algoritma dalam bidang *machine learning* yang digunakan untuk melakukan klasifikasi atau prediksi dengan cara membentuk banyak pohon keputusan (*decision trees*) secara acak. Setiap pohon akan memberikan prediksi, lalu hasil akhir ditentukan berdasarkan suara terbanyak dari pohon (metode voting untuk klasifikasi).

Algoritma ini bekerja dengan data secara acak dari *dataset (bootstrap sampling)* dan membangun setiap pohon dengan hanya menggunakan sebagian fitur yang juga dipilih secara acak. Hal ini membuat random forest sangat tangguh, tidak mudah *overfitting*, dan tetap akurat meskipun data mengandung noise atau tidak seimbang. *Random forest* adalah kumpulan dari banyak pohon keputusan yang saling bekerja sama untuk menghasilkan hasil

prediksi yang lebih akurat dan stabil dibandingkan hanya menggunakan satu pohon saja[3]. Algoritma *Random Forest* dapat dilihat pada Gambar 2.1.



Gambar 2. 1 Algoritma *Random Forest*

Berikut adalah cara kerja *Random Forest* :

1) *Bootstrap Sampling*

Dataset pelatihan di-*sampling* secara acak dengan pengembalian untuk membentuk beberapa subset data (*bootstraps*). Setiap subset digunakan untuk melatih satu pohon keputusan.

2) Pembuatan Pohon Keputusan

Setiap pohon dibangun dari subset data tersebut dengan memilih fitur secara acak di setiap node saat proses pembentukan pohon.

3) *Voting / Averaging*

Untuk tugas klasifikasi, setiap pohon memberikan “*vote*” untuk satu kelas, dan hasil akhir ditentukan berdasarkan suara terbanyak (*majority voting*). Untuk *regresi*, hasilnya adalah rata-rata dari semua output pohon.

2.2.4 Histogram of Oriented Gradient

Histogram of Oriented Gradients (HOG) adalah metode ekstraksi fitur yang banyak digunakan dalam pengolahan citra digital dan *computer vision*, khususnya untuk deteksi objek dan klasifikasi citra. HOG diperkenalkan oleh Dalal dan Triggs (2005) sebagai teknik yang efektif untuk deteksi objek manusia (*human detection*), tetapi sejak itu juga digunakan untuk berbagai tugas pengenalan pola lainnya, termasuk klasifikasi penyakit tanaman berbasis citra daun[8]. Prinsip Dasar HOG adalah menganalisis arah gradien atau tepi pada suatu gambar. Gradien merepresentasikan perubahan intensitas piksel, yang sering kali terjadi di tepi objek atau tekstur. Ciri khas objek dalam gambar dapat direpresentasikan secara efektif melalui distribusi arah gradien di area-area kecil dari citra[9]. HOG memiliki keunggulan yaitu:

- 1) Tahan terhadap perubahan pencahayaan
- 2) Efektif dalam menangkap bentuk dan tepi objek
- 3) Sederhana namun sangat informatif untuk digunakan dalam klasifikasi citra.

Berikut adalah Langkah-langkah umum ekstraksi fitur HOG :

1) *Grayscale Conversion*

Citra RGB diubah menjadi grayscale untuk menyederhanakan perhitungan gradien.

2) *Gradient Computation*

Menghitung gradien horizontal dan vertikal pada setiap piksel dengan operator seperti Sobel.

3) *Cell Division*

Gambar dibagi menjadi blok kecil yang disebut **sel** (misalnya 8×8 piksel).

4) *Orientation Binning*

Dalam setiap sel, arah gradien dihitung dan dimasukkan ke dalam histogram berdasarkan arah (misalnya 0°–180° dibagi ke dalam 9 bin).

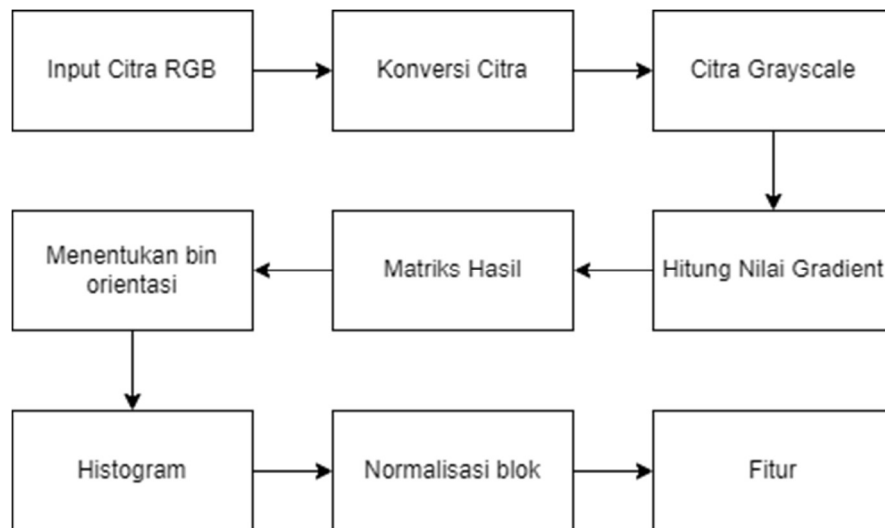
5) *Block Normalization*

Gabungan beberapa sel membentuk blok yang dinormalisasi untuk meningkatkan ketahanan terhadap perubahan pencahayaan dan kontras

6) *Feature Vector*

Semua histogram disatukan menjadi satu vektor fitur yang menggambarkan pola visual dari citra.

Namun ada juga 4 langkah pada metode *Histogram Of Oriented Gradient*[3]. Dapat dilihat pada gambar 2.2.



Gambar 2. 2 Tahapan *HOG*

- 1) Konversi Citra yaitu mengubah citra *RGB* (*Red, Green, Blue*) menjadi citra *grayscale*.
- 2) Perhitungan nilai gradient yang dilakukan untuk setiap piksel pada citra *grayscale* dengan membagi citra menjadi bentuk *cell*.
- 3) Menentukan nilai bin orientasi
- 4) Melakukan *normalisasi* nilai setiap bin orientasi yang kemudian akan menghasilkan fitur

2.2.5 *Confusion Matrix*

Evaluasi bertujuan guna menilai hasil uji coba sistem yang dibuat apakah sudah sesuai antara hasil sistem analisis sentimen dengan hasil yang sebenarnya. *Confusion matrix* adalah tabel yang digunakan untuk menganalisis seberapa baik keakuratan sebuah metode klasifikasi untuk memprediksi kelas suatu data[9]. Guna evaluasi kinerja evaluasi, beberapa parameter yang biasa digunakan yaitu *Precision, Accuracy, Recall, F1-Score*