PENGARUH DATA AUGMENTASI PADA IDENTIFIKASI PENYAKIT DAUN TANAMAN OBAT MENGGUNAKAN K-NEAREST NEIGHBOR

Putu Prianka Vedanty, Made Windu Antara Kesiman, I Made Gede Sunarya
Ilmu Komputer, Universitas Pendidikan Ganesha
Jl. Udayana No.11, Banyuasri, Singaraja, Bali, Indonesia
priankavedanty99@gmail.com

ABSTRAK

Penyakit pada tanaman obat dapat menyebabkan gangguan pertumbuhan, layu, hingga kematian tanaman, sehingga diperlukan identifikasi dini untuk mencegah kerusakan lebih lanjut. Dengan perkembangan teknologi, pengolahan citra digital menjadi solusi yang efisien dalam identifikasi penyakit pada daun tanaman obat. Penelitian menghadapi tantangan berupa keterbatasan data citra penyakit daun tanaman obat yang tidak seimbang (imbalanced data), yang dapat memengaruhi kinerja algoritma klasifikasi. Oleh karena itu, diperlukan proses augmentasi data untuk meningkatkan jumlah dan keragaman data. Penelitian ini bertujuan untuk menganalisis pengaruh augmentasi data terhadap kinerja algoritma K-Nearest Neighbor (K-NN) dalam mengidentifikasi penyakit daun tanaman obat, menggunakan berbagai skenario pengujian pada dataset yang telah diaugmentasi. Penelitian ini melibatkan pengumpulan dataset citra dari beberapa wilayah di Bali seperti wilayah Denpasar Selatan, Kabupaten Tabanan, dan Kabupaten Negara. Dataset diolah menggunakan teknik augmentasi seperti rotasi, resize, dan pembersihan background. Identifikasi dilakukan menggunakan algoritma K-NN, dengan pengujian terhadap enam skenario berdasarkan variasi resolusi dan distribusi data. Evaluasi dilakukan menggunakan confusion matrix untuk menghitung precision, recall, dan F1-score. Proses augmentasi data meningkatkan jumlah dataset dari 300 menjadi 1200 citra. Hasil pengujian menunjukkan skenario kedua, dengan resolusi 700x700 piksel dan dataset balanced, mencapai akurasi tertinggi sebesar 87% pada nilai K=2. Algoritma K-NN menunjukkan performa optimal dengan data yang telah diaugmentasi, menghasilkan prediksi yang lebih akurat dan konsisten.

Kata kunci : *augmentasi, penyakit daun, tanaman obat, k-nearest neighbor*

1. PENDAHULUAN

Penyakit pada tanaman adalah proses yang disebabkan oleh gejala penyakit, dimana proses tersebut akan bekerja setiap harinya dalam jangka waktu yang lama, sehingga mengakibatkan tanaman tidak dapat menjalankan fungsinya dengan baik. Tanaman akan terhambat pertumbuhannya, layu, dan tidak dapat berbuah, bahkan mengakibatkan tanaman menjadi mati apabila terjangkit oleh penyakit tertentu [1].

Organisme pengganggu tanaman (OPT) umumnya disebabkan oleh beberapa faktor diantaranya disebabkan oleh hama, jamur, virus, dan bakteri. Saat ini ilmu pengetahuan berkembang sangat pesat seperti halnya pengolahan citra digital untuk mengidentifikasi penyakit daun tanaman obat. Data diambil dari hasil observasi di beberapa wilayah di Bali. Dengan terbatasnya data yang sangat sedikit, maka dilakukan proses data augmentasi untuk meningkatkan jumlah data dan menjadikan data menjadi lebih beragam. Proses identifikasi dilakukan dengan menggunakan algoritma K-NN. Dilakukan proses data augmentasi bertujuan untuk memperbaiki hasil kinerja dari K-NN dengan melakukan berbagai skenario pengujian dari dataset tanpa menggunakan proses augmentasi dan dataset yang menggunakan proses augmentasi.

Dalam penelitian ini, untuk mendapatkan jumlah dataset citra yang berimbang dilakukan pengumpulan dataset citra dengan menentukan metode akuisisi citra yang tepat. Proses augmentasi data merupakan teknik akusisi yang dilakukan dalam penelitian ini. Augmentasi data adalah proses yang dilakukan secara signifikan untuk memperbanyak jumlah keragaman data tanpa harus mengumpulkan data baru dalam jumlah yang besar [2], dapat dipastikan dengan bertambahnya jumlah data yang dikumpulkan (ketidakseimbangan data), maka kinerja klasifikasi model yang dibangun akan menurun secara signifikan [3].

Maka dari itu perlu dilakukan proses augmentasi, dikarenakan data yang telah dikumpulkan pada penelitian ini jumlahnya sedikit dan tidak berimbang (inbalanced data). Proses augmentasi dilakukan dengan beberapa teknik yaitu teknik resize, cleansing, dan random rotation pada data citra penyakit daun tanaman obat. Resize melibatkan perubahan ukuran gambar menjadi lebih kecil dari ukuran gambar sebelumnya untuk mempermudah dan mempercepat proses pelatihan. Kemudian dilakukan pembersihan background (noise) pada data citra yang akan dirubah menjadi background berwarna hitam. Selanjutnya dilakukan teknik random rotation yang nantinya akan menghasilkan data citra yang telah berotasi berdasarkan derajat yang dipilih menggunakan empat sudut yaitu sudut 0, 90, 180, 270 untuk memperbanyak jumlah data.

Metode K-NN digunakan dalam melakukan identifikasi dikarenakan mampu menanggulangi permasalahan pada bagian karakteristik data yaitu

mampu melakukan pelatihan data dengan efektif dan cepat apabila jumlah data latihnya besar, serta tangguh terhadap dataset yang masih memiliki noise [4], efektif pada data latih yang masih terdapat noise dengan jumlah data latih yang besar [5], klasifikasi sederhana dapat dilakukan dengan menghitung jarak Euclidean untuk mencapai akurasi yang baik berdasarkan jarak terpendek antara data pelatihan dan pengujian.

Adapun beberapa penelitian terkait identifikasi menggunakan algoritma K-NN sebagai berikut. Penelitian yang dilakukan dengan judul penelitian "Diagnosis Hama dan Penyakit Padi Menggunakan Metode Naive Bayes dan K-NN" [6].

Dalam penelitian ini akan dilakukan penggabungan dua metode, yaitu metode Naive Bayes dan K-NN, untuk mengelompokkan kasus (indeks) hama dan penyakit padi, yang terdiri dari 13 spesies hama dan 10 spesies penyakit padi, berdasarkan kelas penyakit. Diagnosis tanpa pengelompokan kasus (tidak diindeks) diterapkan. Hasil pengujian yang diperoleh dengan cross validation adalah nilai rata-rata akurasi pada 153 data uji saat diuji dengan indeks sebesar 92,88%, dan nilai akurasi rata-rata pada 153 data uji saat diuji tanpa indeks menunjukkan 89,63%.

Penelitian selanjutnya dilakukan oleh [7] dengan judul penelitian "Information gain feature Selection untuk klasifikasi penyakit jantung menggunakan kombinasi metode K-NN dan Naive Bayes". Dalam penelitian ini, kami menggabungkan dua metode K-NN dan metode Naive Bayes untuk mengklasifikasikan penyakit jantung dengan memilih fitur perolehan informasi. Pemilihan fitur mengurangi dimensi atribut untuk mendapatkan atribut yang relevan. Dataset yang digunakan terdiri dari 270 buah data yang terdiri dari 13 atribut dan dua label kelas penyakit yaitu dengan penyakit jantung (TPJ) dan tanpa penyakit jantung (TTPJ). Hasil penelitian ini menunjukkan nilai akurasi sebesar 92,31% untuk uji sebaran kelas seimbang menggunakan 6 fitur dengan nilai K=25 dan uji sebaran kelas tidak seimbang menggunakan 4 fitur dengan nilai K=35.

Penelitian selanjutnya dilakukan oleh [8] dengan judul penelitian "Identifikasi Penyakit Daun Jeruk Siam Menggunakan K-NN". Pada penelitian ini kami mengimplementasikan algoritma K-NN dengan menggunakan dataset sebanyak 180 gambar daun jeruk yang dibagi menjadi data latih dan data uji. Penelitian ini dibagi menjadi tiga kelas penyakit. Yaitu 50 gambar daun jeruk sehat sebagai data latih, 50 gambar daun jeruk siam penderita kanker, dan 50 gambar daun jeruk siam penderita penyakit kudis perian. Sebaliknya, kumpulan data-data uji terdiri dari 10 gambar untuk setiap kelas penyakit. Hasil pengujian akhir menghasilkan skor akurasi 70 dengan variasi sesuai K = 21 karena fitur parametrik warna dan tekstur (ASM, entropi, kontras).

Penelitian selanjutnya dilakukan oleh [9] dengan judul penelitian "Identifikasi penyakit daun kentang berdasarkan ciri tekstur dan warna menggunakan metode K-NN". Penelitian ini mengimplementasikan algoritma K-NN untuk mengidentifikasi penyakit daun kentang. Dataset yang digunakan terdiri dari 270 gambar data yang dibagi menjadi tiga kelompok yaitu 90 gambar data latih, 90 gambar data uji, dan 90 gambar data validasi. Hasil pengujian akhir menghasilkan skor akurasi sebesar 80 dengan nilai K = 3.

Berikut penelitian yang dilakukan oleh [10] dengan judul penelitian: "Analisa Pengaruh Data Augmentasi Pada Klasifikasi Bumbu Dapur Menggunakan Convolutional Neural Network." Dalam studi ini, kami menerapkan augmentasi kumpulan data kendaraan standar menggunakan pendekatan pemangkasan acak, rotasi, dan perturbasi.

2. TINJAUAN PUSTAKA

2.1. K-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah teknik yang menggunakan perhitungan yang diatur. Perhitungan ini adalah prosedur learning. K-Nearest Neighbor K-NN selesai dengan mencari banyak k artikel dalam informasi persiapan itu terdekat (seperti) artikel di informasi baru atau pengujian informasi [11].

Metode KNN diterapkan untuk mengatasi permasalahan identifikasi yang diukur secara kualitatif maupun kuantitatif. Sebelum mencari jarak data ke tetangga adalah menentukan nilai K tetangga (neighbor) [12].

Untuk mendefinisikan jarak antara dua titik pada data training dan titik pada data testing, yaitu menggunakan rumus Euclidean Distance. Hasil dari perhitungan jarak, kemudian diurutkan dari data dengan nilai terkecil hingga terbesar sesuai dengan nilai k yang telah ditentukan sebelumnya.

2.2. Augmentasi Data

Augmentasi data adalah proses memodifikasi atau memanipulasi suatu citra sehingga citra asli dalam bentuk yang telah disiapkan berubah bentuk dan posisinya. Augmentasi data bertujuan untuk memungkinkan mesin belajar dan mengenali dari gambar yang berbeda, dan dapat digunakan untuk merekonstruksi data [13].

2.3. Pengolahan Citra

Representasi suatu objek dalam dua dimensi dikenal sebagai citra. Citra ini dapat direkam menggunakan perangkat optik yang menangkap pantulan cahaya dari objek tiga dimensi; misalnya, kamera dapat mengambil objek tiga dimensi dan memprosesnya untuk menghasilkan citra dua dimensi. Citra digital terdiri dari elemen yang dikenal sebagai piksel atau nilai tingkat abu-abu, dengan setiap piksel diberi lokasi dan nilai tertentu. Intensitas piksel diskrit ini berkisar dari 0 (hitam) hingga 255 (putih) [14].

2.4. Confussion Matrix

Confusion matrix adalah tabel yang menggambarkan hasil klasifikasi, menampilkan

jumlah data uji yang diklasifikasikan dengan benar dan jumlah data yang salah klasifikasi. Matriks ini mencakup empat elemen utama: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN), yang digunakan untuk mengevaluasi performa model klasifikasi [15].

2.5. Precision

Precision mengukur proporsi prediksi benar dalam suatu kelas dibandingkan dengan seluruh prediksi untuk kelas tersebut. Precision penting untuk menilai akurasi klasifikasi, terutama dalam memastikan bahwa prediksi penyakit daun tidak mengandung terlalu banyak kesalahan deteksi yang salah (false positives) [15].

2.6. Recall

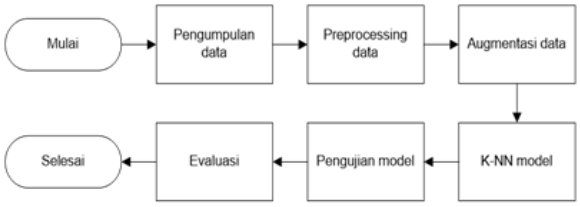
Recall menunjukkan seberapa baik model dapat mendeteksi seluruh sampel yang sebenarnya berada dalam suatu kelas. Metode ini penting untuk mengevaluasi kemampuan sistem dalam mengidentifikasi semua daun yang terinfeksi dalam setiap kelas penyakit, karena memberikan informasi tentang tingkat keberhasilan model dalam mengurangi kesalahan deteksi (false negatives) [15].

2.7. F1-Score

F1-Score merupakan gabungan precision dan recall, memberikan gambaran keseimbangan antara keduanya. F1-score penting karena mempertimbangkan situasi di mana precision atau recall lebih rendah, memberikan nilai rata-rata harmonis yang menunjukkan efisiensi keseluruhan model dalam mengklasifikasikan penyakit daun [15].

3. METODE PENELITIAN

Adapun tahapan-tahapan penelitian yang dilakukan. Berikut merupakan beberapa tahapan yang dilakukan untuk mengidentifikasi penyakit daun tanaman obat sebagai berikut:



Gambar 1. Metode Penelitian

Dari Tabel 1 dapat dilihat bahwa tahapan penelitian dimulai dari tahap pengumpulan data citra penyakit daun tanaman obat. Selanjutnya dilakukan preprocessing data pada citra dataset seperti proses resize dan cleansing data. Dilanjutkan dengan proses augmentasi data, melakukan penentuan model K-NN, pengujian model, dan terakhir tahap evaluasi hasil. Adapun pembahasan lebih lengkap sebagai berikut

3.1. Dataset

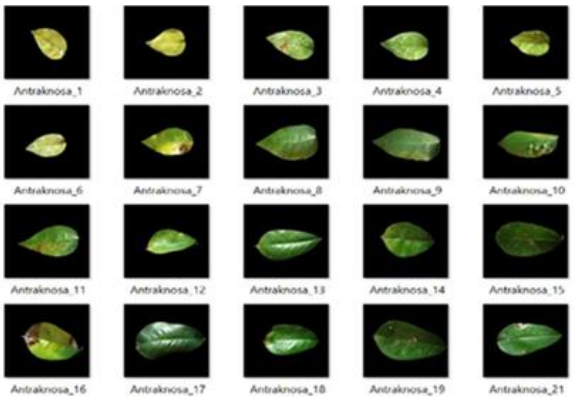
Pada penelitian ini dataset yang digunakan didapat berdasarkan hasil dari obervasi di beberapa wilayah di Bali seperti wilayah Denpasar Selatan, Kabupaten Tabanan, dan Kabupaten Negara. Total dataset yang diperoleh yaitu berjumlah 300 data citra yang terdiri dari 15 kelas penyakit tanaman obat yaitu penyakit antraknosa, bercak fusarium, bercak glosporium, bercak daun, cendawan daun, embun tepung, jamur akar putih, jamur upas, kaker rating, karat merah, layu bakteri, layu fusarium, virus cpvd, virus kuning, virus mozaik yang sebelumnya sudah dilakukan wawancara dengan pakar penyakit tanaman.

3.2. Preprocessing Dataset

Setelah dilakukan pengumpulan dataset, langkah selanjutnya yaitu melakukan preprocessing data. Adapun beberapa tahapan yang dilakukan yaitu sebagai berikut:

3.2.1. Pembagian Dataset

Setelah dataset telah dikumpulkan, lalu dikelompokkan menjadi 15 kelas penyakit tanaman obat yaitu penyakit antraknosa, bercak fusarium, bercak glosporium, bercak daun, cendawan daun, embun tepung, jamur akar putih, jamur upas, kaker rating, karat merah, layu bakteri, layu fusarium, virus cpvd, virus kuning, virus mozaik. Data awal yang berhasil dikumpulkan yaitu berjumlah 300 data citra. Kemudian dibagi kedalam data pelatihan menjadi 225 data citra dan kedalam data pengujian menjadi 75 data citra. Berikut merupakan contoh dataset yang telah dikumpulkan sebagai berikut:



Gambar 2. Sampel Dataset

Untuk memaksimalkan hasil akurasi yang didapatkan saat pengujian, diperlukan jumlah dataset Maka dilakukan proses augmentasi manual dengan melakukan rotasi 180 derajat. Setelah dilakukan augmentasi data citra bertambah menjadi 1200 data citra. Kemudian dibagi kedalam data pelatihan menjadi 900 data citra dan kedalam data pengujian menjadi 300 data citra. yang besar dan beragam. Maka dilakukan proses augmentasi manual dengan melakukan rotasi 180 derajat. Setelah dilakukan augmentasi data citra bertambah menjadi 1200 data

citra. Kemudian dibagi kedalam data pelatihan menjadi 900 data citra dan kedalam data pengujian menjadi 300 data citra.

3.2.2. Resize Dataset

Gambar 3. Sampel Dataset Resize

Data yang sudah dikumpulkan sebelumnya memiliki ukuran resolusi yang berbeda beda, maka dilakukan proses resize menjadi beberapa ukuran yaitu ukuran 900x900 pixel, 700x700 pixel, dan 500x500 pixel untuk dilakukan beberapa skenario pengujian.

3.3. Augmentasi Data

Gambar 4. Sampel Dataset Augmentasi

Augmentasi merupakan proses yang dilakukan bertujuan untuk menambah jumlah data tanpa benar-benar mengumpulkan data baru dalam jumlah yang besar dan beragam. Dikarenakan dataset yang telah dikumpulkan masih sedikit. Maka dilakukan proses augmentasi dengan melakukan rotasi secara random yaitu dari 0 derajat sampai dengan 180 derajat.

3.4. K-NN Model

Sebelum dilakukan proses klasifikasi dengan algoritma K-NN, pada penelitian ini akan dilakukan klasifikasi dengan menerapkan proses ekstraksi fitur. Fitur yang digunakan yaitu fitur warna dan tekstur. Hasil ekstraksi fitur warna yang akan digunakan terdiri dari 14 parameter atribut yaitu mencari Nilai Rerata Citra (RGB) mencari nilai Standar Deviasi Citra (RGB), mencari Nilai Rerata Citra Hue, Saturation, Value (HSV), mencari Nilai Standar Deviasi Citra Hue, Saturation, Value (HSV), dan mencari Nilai Rerata Citra Grayscale dan Standar Deviasi Citra Grayscale. Berikut merupakan tabel hasil ekstraksi fitur warna sebagai berikut:

Tabel 1. Hasil Ekstraksi Fitur Warna

Mean B	Mean G	Mean R	Std B	Std G	Std R
9.400	21.860	22.091	27.742	59.679	60.248
8.925	22.002	21.992	26.642	61.616	61.499
14.317	26.906	23.368	36.306	64.461	56.717
12.700	26.125	21.049	32.802	63.259	51.680
9.400	21.860	22.091	27.742	59.679	60.248

Kemudian untuk hasil ekstraksi fitur tekstur yang akan digunakan terdiri dari 7 parameter atribut yaitu mencari nilai Contrast, Correlation, Energy, Homogeneity, Entropy, Dissimilarity, ASM. Berikut merupakan tabel hasil ekstraksi fitur tekstur sebagai berikut:

Tabel 2. Hasil Ekstraksi Fitur Tekstur

Contrast	Correlation	Energy
15.265	0.997	0.873
13.173	0.998	0.878
15.498	0.998	0.840
15.764	0.997	0.842
13.204	0.997	0.877

Selanjutnya untuk melakukan identifikasi pada algoritma K-NN, tahap awal yang dilakukan adalah menentukan himpunan K yang akan menjadi kromosom atau calon K paling optimal yang akan ditemukan. Nilai himpunan K yang digunakan yakni: { 2, 4, 6, 8, 10 }. Tahap selanjutnya adalah menerapkan metode K-NN dengan masing-masing nilai kromosom dalam himpunan K. Kemudian dihitung tingkat akurasi dengan menggunakan metode Confusion Matrix, jika nilai yang didapatkan optimal, maka akan dijadikan sebagai solusi optimal, jika belum maka proses akan diulang dengan himpanan K lainnya.

4. HASIL DAN PEMBAHASAN

4.1. Pengujian Model

Tahapan selanjutnya yaitu melakukan identifikasi dengan pengujian model yang telah ditentukan sebelumnya. Pada bagian ini juga akan dijelaskan hasil dari penelitian yang telah dilakukan mulai dari hasil Confusion Matrix, hasil pengujian metode K-NN dengan 6 skenario pengujian pada dataset.

Nilai (K)	Presisi(%)	Recall(%)	F1-Score(%)
2	85,79	85,67	85,32
4	84,35	84,33	83,91
6	79,12	78,67	77,95
8	73,44	72,67	72,03
10	74,05	72	71,15

Gambar 5. Hasil Iterasi Nilai K pada Skenario Pengujian Pertama

Pada skenario pertama akan dilakukan pengujian pada dataset 900x900 pixel dengan jumlah 300 data pengujian balanced dengan background hitam bersih dengan iterasi nilai K yang sudah ditentukan sebelumnya.

Nilai (K)	Presisi(%)	Recall(%)	F1-Score(%)
2	87,79	87	87,01
4	80,36	80,67	80,6
6	79,57	79	78,18
8	74,44	73,33	72,50
10	72,90	71,33	70,37

Gambar 6. Hasil Iterasi Nilai K pada Skenario Pengujian Kedua

Pada skenario kedua akan dilakukan pengujian pada dataset 700x700 pixel dengan jumlah 300 data pengujian balanced dengan background hitam bersih dengan iterasi nilai K yang sudah ditentukan sebelumnya.

Nilai (K)	Presisi(%)	Recall(%)	F1-Score(%)
2	84,08	83	82,87
4	78,59	78,67	77,97
6	75,05	75	74,04
8	70,16	70	68,82
10	65,90	66	64,71

Gambar 7. Hasil Iterasi Nilai K pada Skenario Pengujian Ketiga

Pada skenario ketiga akan dilakukan pengujian pada dataset 500x500 pixel dengan jumlah 300 data pengujian balanced dengan background hitam bersih dengan iterasi nilai K yang sudah ditentukan sebelumnya.

Nilai (K)	Presisi(%)	Recall(%)	F1-Score(%)
2	83,14	81,67	81,11
4	79,21	78,98	78,51
6	75,39	74,66	74,09
8	73,61	73,32	72,65
10	71,02	70,89	70,01

Gambar 8. Hasil Iterasi Nilai K pada Skenario Pengujian Keempat

Pada skenario keempat akan dilakukan pengujian pada dataset 900x900 pixel dengan jumlah 371 data pengujian inbalanced dengan background hitam bersih dengan iterasi nilai K yang sudah ditentukan sebelumnya.

Nilai (K)	Presisi(%)	Recall(%)	F1-Score(%)
2	86,07	84,64	84,47
4	75,97	75,20	74,25
6	75,95	74,66	73,84
8	67,50	67,92	66,22
10	66,63	67,12	64,98

Gambar 9. Hasil Iterasi Nilai K pada Skenario Pengujian Kelima

Pada skenario kelima akan dilakukan pengujian pada dataset 700x700 pixel dengan jumlah data 371 data test inbalanced dengan background hitam bersih dengan iterasi nilai K yang sudah ditentukan sebelumnya.

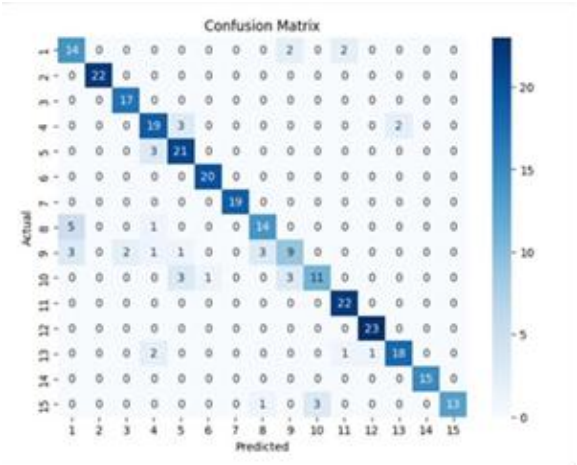
Nilai (K)	Presisi(%)	Recall(%)	F1-Score(%)
2	82,23	79,51	78,84
4	74,84	74,39	73,76
6	72,42	71,16	70,48
8	69,86	68,73	67,32
10	68,41	67,39	66,15

Gambar 10. Hasil Iterasi Nilai K pada Skenario Pengujian Keenam

Pada skenario terakhir yaitu keenam akan dilakukan pengujian pada dataset 500x500 pixel dengan jumlah 371 data pengujian inbalanced dengan background hitam bersih dengan iterasi nilai K yang sudah ditentukan sebelumnya.

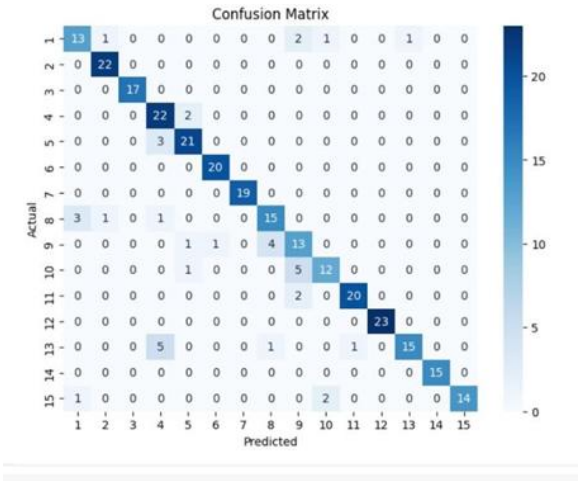
4.2. Evaluasi Model

Untuk menguji kinerja dibutuhkan tools atau alat yang dapat mengukur kinerja pemodelan yang digunakan. Dalam klasifikasi, salah satu alat yang dapat digunakan untuk mengukur kinerja klasifikasi adalah dengan menghitung Confusion Matrix. Berikut merupakan gambar hasil Confusion Matrix berdasarkan skenario pengujian sebagai berikut:



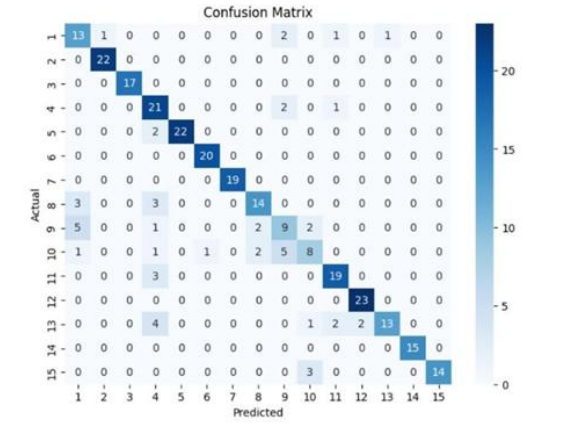
Gambar 11. Confusion Matrix Skenario Pengujian Pertama

Pada gambar 11 merupakan hasil Confusion Matrix skenario pengujian pertama pada nilai K=2 sebagai sampel dengan menerapkan augmentasi data pada dataset ukuran 900x900 pixel jumlah 300 data pengujian balanced dengan background hitam bersih. Dari gambar tersebut dapat dianalisis bahwa data yang diprediksi benar paling banyak yaitu sebanyak 23 data pada kelas penyakit 12.



Gambar 12. Confusion Matrix Skenario Pengujian Kedua

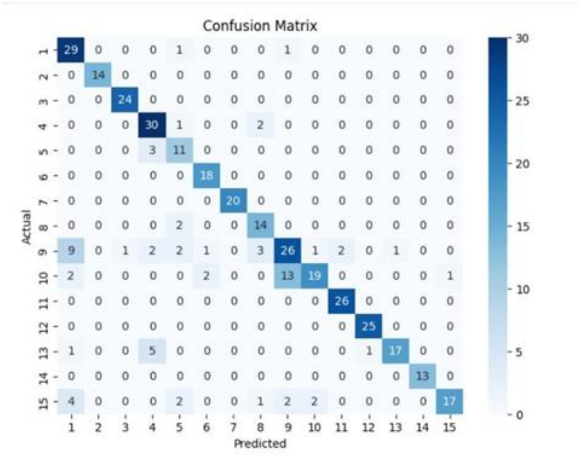
Pada gambar 12 merupakan hasil Confusion Matrix dari skenario pengujian kedua pada nilai K=2 sebagai sampel dengan menerapkan augmentasi data pada dataset ukuran 700x700 pixel jumlah 300 data pengujian balanced dengan background hitam bersih. Dari gambar tersebut dapat dianalisis bahwa data yang diprediksi benar paling banyak yaitu sebanyak 23 data pada kelas penyakit 12.



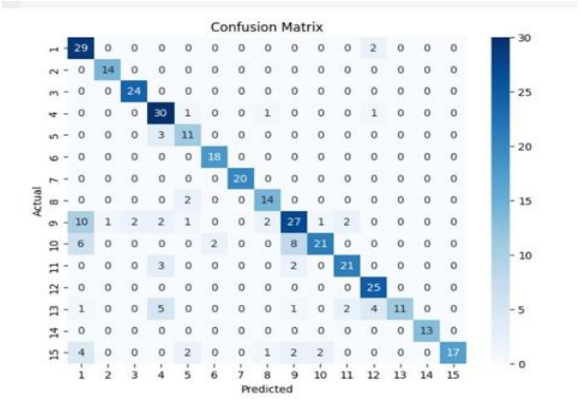
Gambar 13. Confusion Matrix Skenario Pengujian Ketiga

Pada gambar 13 merupakan hasil Confusion Matrix dari skenario pengujian ketiga pada nilai K=2 sebagai sampel dengan menerapkan augmentasi data pada dataset ukuran 700x700 pixel jumlah 300 data pengujian balanced dengan background hitam bersih. Dari gambar tersebut dapat dianalisis bahwa data yang diprediksi benar paling banyak yaitu sebanyak 23 data pada kelas penyakit 12.

Pada gambar 14 merupakan hasil Confusion Matrix dari skenario pengujian keempat pada nilai K=2 sebagai sampel dengan menerapkan augmentasi data pada dataset ukuran 900x900 pixel jumlah 371 data pengujian inbalanced dengan background hitam bersih. Dari gambar tersebut dapat dianalisis bahwa data yang diprediksi benar paling banyak yaitu sebanyak 30 data pada kelas penyakit 4.

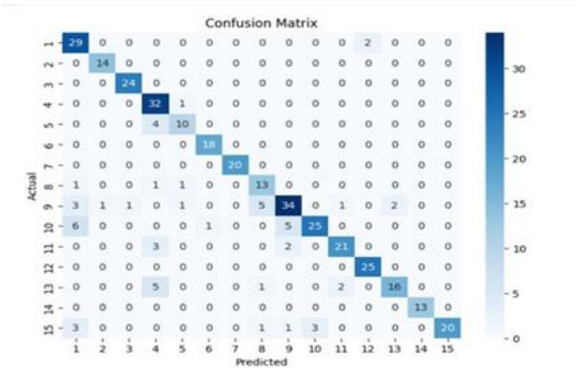


Gambar 14. Confusion Matrix Skenario Pengujian Keempat



Gambar 15. Confusion Matrix Skenario Pengujian Kelima

Pada gambar 15 merupakan hasil Confusion Matrix dari skenario pengujian kelima pada nilai K=2 sebagai sampel dengan menerapkan augmentasi data pada dataset ukuran 700x700 pixel jumlah 371 data pengujian inbalanced dengan background hitam bersih. Dari gambar tersebut dapat dianalisis bahwa data yang diprediksi benar paling banyak yaitu sebanyak 34 data pada kelas penyakit 9.



Gambar 16. Confusion Matrix Skenario Pengujian Keenam

Pada gambar 16 merupakan hasil Confusion Matrix dari skenario pengujian keenam pada nilai K=2 sebagai sampel dengan menerapkan augmentasi data pada dataset ukuran 500x500 pixel jumlah 371 data pengujian inbalanced dengan background hitam bersih. Dari gambar tersebut dapat dianalisis bahwa data yang diprediksi benar paling banyak yaitu sebanyak 30 data pada kelas penyakit 3.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian yang telah dilakukan proses identifikasi dengan algoritma K-NN dengan menerapkan 6 skenario pengujian, skenario pengujian ke 2 mendapatkan nilai akurasi lebih baik saat melakukan proses pengujian yaitu sebesar 87% pada dataset 700x700 pixel dengan jumlah 300 data pengujian balanced yang sudah dilakukan proses augmentasi dengan nilai iterasi K=2.

DAFTAR PUSTAKA

[1] Rahmawati Reny, *Cepat Dan Tepat Berantas Hama Dan Penyakit Tanaman*. Yogyakarta: Pustaka Baru Press, 2021.

[2] J. Sanjaya And M. Ayub, "Augmentasi Data Pengenalan Citra Mobil Menggunakan Pendekatan Random Crop, Rotate, Dan Mixup," *Jurnal Teknik Informatika Dan Sistem Informasi*, Vol. 6, No. 2, Aug. 2020, Doi: 10.28932/Jutisi.V6i2.2688.

[3] M. Resa Arif Yudianto, P. Sukmasetya, R. Abul Hasani, And D. Sasongko, "Pengaruh Data Preprocessing Terhadap Imbalanced Dataset Pada Klasifikasi Citra Sampah Menggunakan Algoritma Convolutional Neural Network," *Building Of Informatics, Technology And Science (Bits)*, Vol. 4, No. 3, Dec. 2022, Doi: 10.47065/Bits.V4i3.2575.

[4] A. P. Ayudhitama And U. Pujianto, "Analisa 4 Algoritma Dalam Klasifikasi Penyakit Liver Menggunakan Rapidminer," 2020.

[5] R. Enggar Pawening, W. Ja, And Far Shudiq, "Klasifikasi Kualitas Jeruk Lokal Berdasarkan Tekstur Dan Bentuk Menggunakan Metode K-Nearest Neighbor (K-Nn)," 2020. [Online]. Available: Http://Ejournal.Unuja.Ac.Id/Index.Php/Core

[6] R. M. Bianome, Y. Y. Nabuasa, And D. R. Sina, "Diagnosa Hama Dan Penyakit Pada Tanaman Padi Menggunakan Metode Naive Bayes Dan K-Nearest Neighbor," *Jurnal Komputer Dan Informatika*, Vol. 8, No. 2, Pp. 156–162, Oct. 2020, Doi: 10.35508/Jicon.V8i2.2906.

[7] S. Hidayatul, A. Aini, Y. A. Sari, And A. Arwan, "Seleksi Fitur Information Gain Untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor Dan Naïve Bayes," 2018. [Online]. Available: Http://J-Ptiik.Ub.Ac.Id

[8] R. H. Ariesdianto, Z. E. Fitri, A. Madjid, And A. M. N. Imron, "Identifikasi Penyakit Daun Jeruk Siam Menggunakan K-Nearest Neighbor," *Jurnal Ilmu Komputer Dan Informatika*, Vol. 1, No. 2, Pp. 133–140, Nov. 2021, Doi: 10.54082/Jiki.14.

[9] A. M. Lesmana, R. P. Fadhillah, And C. Rozikin, "Identifikasi Penyakit Pada Citra Daun Kentang Menggunakan Convolutional Neural Network (Cnn)," *Jurnal Sains Dan Informatika*, Vol. 8, No. 1, Pp. 21–30, Jun. 2022, Doi: 10.34128/Jsi.V8i1.377.

[10] W. M. Pradnya D And A. P. Kusumaningtyas, "Analisis Pengaruh Data Augmentasi Pada Klasifikasi Bumbu Dapur Menggunakan Convolutional Neural Network," *Jurnal Media Informatika Budidarma*, Vol. 6, No. 4, P. 2022, Oct. 2022, Doi: 10.30865/Mib.V6i4.4201.

[11] W. Saputro, D. B. Sumantri, S. Tinggi, I. Komputer, And C. Karya Informatika, "Implementasi Citra Digital Dalam Klasifikasi Jenis Buah Anggur Dengan Algoritma K-Nearest Neighbors (Knn) Dan Data Augmentasi Digital Image Implementation In Grape Classification With K-Nearest Neighbors (Knn) Algorithm And Augmentation Data," *Journal Of Information Technology And Computer Science (Intecoms)*, Vol. 5, No. 2, 2022.

[12] A. Yhurinda, P. Putri, A. Sodik, I. T. Adhi, And T. Suarabaya, "Identifikasi Penyakit Tanaman Kopi Arabika Dengan Metode K-Nearest Neighbor (K-Nn)," 2019.

[13] P. M. Widodo, A. Jazuli, And E. Wijayanti, "Teknik Pengolahan Citra Digital Untuk Identifikasi Penyakit Pada Daun Tanaman Kopi," *Jurnal Dialektika Informatika (Detika)*, Vol. 5, No. 1, Pp. 39–44, 2024, Doi: 10.24176/Detika.V5i1.13943.

[14] P. M. Widodo, A. Jazuli, And E. Wijayanti, "Teknik Pengolahan Citra Digital Untuk Identifikasi Penyakit Pada Daun Tanaman Kopi," *Jurnal Dialektika Informatika (Detika)*, Vol. 5, No. 1, Pp. 39–44, 2024, Doi: 10.24176/Detika.V5i1.13943.

[15] E. Safitri *Et Al.*, "Klasifikasi Penyakit Daun Anggur Berbasis Citra Menggunakan Metode K-Nearest Neighbors (Knn)," 2024.