

KLASIFIKASI DAUN TEBU MENGGUNAKAN PENDEKATAN BERBASIS GRAYSCALE, PCA DAN SVM

Hengky Wisianto ¹⁾ Hizkia Rahuel Anglie ²⁾ Alfian Effendi ³⁾

^{1) 2) 3)} Teknik Informatika Universitas Tarumanagara

Jl. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

email : hengky.535230094@stu.untar.ac.id ¹⁾ hizkia.535230073@stu.untar.ac.id ²⁾ alfian.535230089@stu.untar.ac.id ³⁾

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem klasifikasi otomatis terhadap empat kategori kondisi daun tebu yaitu *healthy*, *rust*, *redrot*, dan *yellow* menggunakan dataset sebanyak 800 citra. Metodologi dimulai dengan pra-pemrosesan berupa *resizing* dan konversi *grayscale*. Untuk mengatasi dimensi fitur yang tinggi, digunakan *Principal Component Analysis (PCA)* untuk mereduksi dimensi dengan mempertahankan 98% variansi. Data hasil reduksi kemudian diklasifikasikan menggunakan *Support Vector Machine (SVM)* kernel linear. Hasil eksperimen menunjukkan akurasi tertinggi diperoleh saat pelatihan menggunakan 80% data dengan nilai akurasi 81,76%. Namun demikian, kategori tertentu seperti *healthy* dan *yellow* mengalami penurunan akurasi yang signifikan. Kelemahan utama dari model ini adalah hilangnya informasi warna akibat konversi *grayscale*. Oleh karena itu, meskipun model menunjukkan performa baik dalam kasus tertentu, pendekatan ini memiliki keterbatasan dalam membedakan gejala penyakit berbasis warna.

Key words

Grayscale, Principal Component Analysis (PCA), Support Vector Machine (SVM)

1. Pendahuluan

Di Indonesia, tebu adalah komoditas perkebunan yang esensial untuk produksi gula nasional. Kualitas gula sangat bergantung pada kualitas tebu yang digunakan. Oleh karena itu, tantangan seperti serangan mikroorganisme yang dapat merusak mutu tebu menjadi ancaman serius bagi industri ini [1].

Apabila serangan mikroorganisme tersebut tidak cepat di deteksi atau dipisahkan maka dapat mengurangi kualitas tebu yang akan di panen, dan juga metode pemisahan tebu secara manual dapat mengurangi hasil panen dan dapat mengurangi produktivitas, sehingga diperlukan sebuah teknologi pendeteksian tebu untuk meningkatkan hasil panen tebu [2].

Seiring dengan kemajuan teknologi, pengolahan citra digital dan kecerdasan buatan menawarkan solusi yang menjanjikan untuk mengatasi permasalahan tersebut. Sistem identifikasi otomatis berbasis citra dapat

memberikan diagnosis yang lebih cepat, objektif, dan konsisten. Penelitian ini mengusulkan sebuah sistem klasifikasi kondisi daun tebu dengan memanfaatkan serangkaian metode komputasi yang efisien.

Pendekatan utama dalam penelitian ini dimulai dengan konversi citra daun tebu menjadi format *grayscale*. Langkah ini bertujuan untuk menyederhanakan data citra dengan menghilangkan informasi warna dan memfokuskan analisis pada fitur-fitur esensial seperti tekstur, bentuk, dan pola gradasi keabuan yang sering menjadi indikator kunci kesehatan daun [3]. Selanjutnya, untuk mengekstraksi fitur yang paling signifikan dari citra *grayscale* dan mereduksi dimensi data, digunakan metode *Principal Component Analysis (PCA)*. PCA secara efektif mentransformasi data ke dalam komponen utama yang paling banyak merepresentasikan variasi data, sehingga proses komputasi menjadi lebih ringan dan terfokus pada informasi yang relevan [4].

Fitur yang telah diekstraksi oleh PCA kemudian akan menjadi masukan bagi algoritma klasifikasi *Support Vector Machine (SVM)*. SVM dipilih karena kemampuannya yang telah terbukti andal dalam menangani masalah klasifikasi dengan data berdimensi tinggi dan kemampuannya menemukan *hyperplane* pemisah terbaik antar kelas [5].

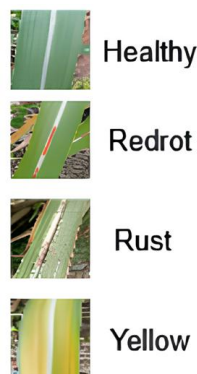
Dengan mengintegrasikan pendekatan berbasis *grayscale*, PCA, dan SVM, penelitian ini bertujuan untuk membangun sebuah model yang akurat dan efisien. Hasil dari penelitian ini diharapkan dapat menjadi pengembangan sistem guna mendukung peningkatan dan kualitas hasil pertanian daun tebu.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode klasifikasi citra digital. Tujuan utamanya adalah menguji efektivitas kombinasi teknik *Principal Component Analysis (PCA)* dan *Support Vector Machine (SVM)* dalam mengklasifikasikan kondisi daun tebu berdasarkan dataset citra.

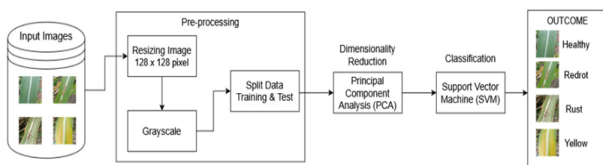
2.1. Dataset

Data yang digunakan dalam penelitian ini adalah citra daun tebu yang telah dikategorikan sesuai jenis penyakitnya. Sumber data diperoleh dari situs Kaggle melalui dataset “*Sugarcane Leaf Disease*”, yang terdiri dari empat kategori utama, yaitu *healthy*, *rust*, *redrot*, dan *yellow* (Gambar 1). Seluruh data disimpan dalam format jpg dan telah dikelompokkan ke dalam satu direktori sesuai label penyakit. Dari keseluruhan populasi yang tersedia, peneliti menggunakan sebanyak 800 citra sebagai sampel yang diolah dan dianalisis lebih lanjut dalam penelitian ini.



Gambar 1. Kategori Daun Tebu

2.2. Diagram Alir Penelitian



Gambar 2. Diagram Alir Proses Penelitian

Diagram alir dalam penelitian ini menggambarkan alur proses klasifikasi citra daun tebu mulai dari pengumpulan data hingga evaluasi hasil klasifikasi. Tahapan awal dimulai dengan pengumpulan dataset citra daun tebu dari sumber terbuka (Kaggle). Data ini kemudian diproses dalam tahap pra-pemrosesan, yang meliputi perubahan ukuran citra menjadi 128×128 piksel dan konversi warna ke dalam bentuk *grayscale*. Tahap ini bertujuan untuk menyamakan resolusi serta menyederhanakan representasi visual dari citra [6].

Setelah citra mengalami pra-pemrosesan, data tersebut dimasukkan ke tahap ekstraksi fitur menggunakan *Principal Component Analysis* (PCA). PCA digunakan untuk mereduksi jumlah fitur dari ribuan piksel menjadi sejumlah kecil komponen utama yang tetap merepresentasikan informasi penting dalam citra. Reduksi ini bertujuan untuk mengurangi beban komputasi dan mencegah *overfitting* pada model klasifikasi [7].

Fitur hasil reduksi PCA kemudian dilanjutkan ke proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Model SVM dilatih menggunakan data

latih, kemudian diuji menggunakan data uji dalam beberapa rasio pembagian data (60:40, 70:30, dan 80:20). Proses klasifikasi menghasilkan prediksi kelas untuk masing-masing citra.

Tahap akhir adalah evaluasi model klasifikasi, di mana hasil prediksi dibandingkan dengan label aktual untuk menghasilkan matrik evaluasi berupa *confusion matrix*, akurasi, *precision*, *recall*, dan *F1-score*. Evaluasi ini bertujuan untuk mengukur performa model dalam mengenali dan membedakan kategori penyakit daun tebu secara otomatis [8].

Secara keseluruhan, diagram alir mencerminkan alur logis dari proses transformasi citra mentah hingga menjadi keputusan klasifikasi yang dapat membantu dalam deteksi penyakit tanaman tebu secara digital.

2.3. Teknik Pengolahan dan Analisis data

Proses pengolahan dan analisis data dimulai dari tahap pra-pemrosesan, yaitu mengubah ukuran citra menjadi 128×128 piksel dan mengkonversinya ke format *grayscale* untuk menyederhanakan komputasi. Selanjutnya dilakukan reduksi dimensi menggunakan PCA, dengan mempertahankan 98% variansi data untuk menghasilkan fitur utama. Fitur yang telah direduksi kemudian digunakan sebagai *input* dalam algoritma SVM untuk membangun model klasifikasi. Evaluasi model dilakukan menggunakan *confusion matrix* serta dihitung metrik seperti akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja model secara menyeluruh.

2.4. Alat Penelitian

Alat yang digunakan dalam penelitian ini terdiri dari perangkat keras dan perangkat lunak. Perangkat keras meliputi laptop dengan spesifikasi minimum RAM 8GB dan prosesor Intel Core i5. Perangkat lunak yang digunakan adalah MATLAB, yang dimanfaatkan untuk melakukan pemrosesan citra, ekstraksi fitur menggunakan PCA, serta implementasi algoritma klasifikasi SVM. Dataset citra daun tebu juga merupakan bagian dari instrumen utama karena menjadi objek pengujian dalam eksperimen.

2.5. Algoritma

2.5.1. Grayscale

Grayscale merupakan metode representasi citra digital dalam bentuk tingkat keabuan yang hanya memiliki satu kanal warna, yaitu intensitas cahaya [9]. Metode ini banyak digunakan dalam pengolahan citra karena mampu menyederhanakan data dari citra berwarna (RGB) menjadi citra yang hanya mengandung informasi kecerahan, tanpa informasi warna. Dengan demikian, penggunaan *grayscale* dapat mempercepat proses komputasi dan mengurangi kompleksitas fitur yang akan digunakan pada tahap selanjutnya seperti ekstraksi ciri atau klasifikasi.

Dalam konteks citra RGB, setiap piksel memiliki tiga komponen warna utama: merah (*Red*), hijau (*Green*), dan biru (*Blue*). Untuk mengubah citra RGB menjadi *grayscale*, digunakan rumus pembobotan luminansi berdasarkan sensitivitas mata manusia terhadap masing-masing warna. Rumus konversi tersebut adalah sebagai berikut:

$$I_{\text{grayscale}} = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1) [9]$$

Keterangan:

R = nilai intensitas kanal merah (*Red*)

G = nilai intensitas kanal hijau (*Green*)

B = nilai intensitas kanal biru (*Blue*)

Pembobotan ini menunjukkan bahwa mata manusia lebih sensitif terhadap warna hijau, sehingga nilai G diberi bobot lebih tinggi dibandingkan R dan B.

Dalam penelitian ini, seluruh citra daun tebu dikonversi menjadi *grayscale* sebelum masuk ke proses ekstraksi fitur menggunakan PCA. Penggunaan *grayscale* bertujuan untuk mengurangi dimensi data (dari 3 kanal menjadi 1 kanal) serta mengurangi pengaruh variasi warna yang dapat memperumit proses klasifikasi.

2.5.2. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah metode statistik yang digunakan untuk mereduksi dimensi data tanpa kehilangan informasi penting secara signifikan [10]. PCA bekerja dengan mencari transformasi linier yang memetakan data asli ke sistem koordinat baru berdasarkan arah variansi terbesar dalam data. Proses ini menghasilkan sejumlah komponen utama (*principal components*) yang merupakan kombinasi linier dari fitur asli.

Langkah-langkah PCA meliputi [10]:

1. Menghitung nilai rata-rata setiap fitur.
2. Menyusun matriks kovarians dari data.
3. Menghitung nilai *eigen value* dan *eigen vector* dari matriks kovarians.
4. Memilih *eigen vector* dengan nilai *eigen* terbesar.
5. Memproyeksikan data ke ruang dimensi baru menggunakan *eigen vector* yang terpilih.

Persamaan umum PCA dalam bentuk sistem persamaan linear dituliskan sebagai:

$$C_v = \lambda_v \quad (2) [10]$$

Keterangan:

C = matriks kovarians

v = *eigen vector*

λ = *eigen value*

Pada penelitian ini, PCA digunakan untuk mereduksi 16.384 fitur piksel menjadi 296 - 381 komponen utama, tergantung pada proporsi data latih yang digunakan, sambil mempertahankan 98% variansi total.

2.5.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma klasifikasi berbasis margin yang bekerja dengan mencari *hyperplane* terbaik yang dapat memisahkan data ke dalam dua atau lebih kelas secara optimal [11]. Dalam ruang berdimensi tinggi, SVM membagi data menggunakan batas yang memaksimalkan jarak (*margin*) antara dua kelas terdekat. Semakin besar margin yang diperoleh, semakin baik generalisasi model terhadap data baru.

SVM digunakan dalam penelitian ini karena kemampuannya yang tinggi dalam menangani data berdimensi besar hasil dari ekstraksi PCA, serta keandalannya dalam menghindari *overfitting*. Algoritma SVM digunakan dalam tiga skenario rasio pembagian data: 60:40, 70:30, dan 80:20 antara data latih dan data uji. Kernel yang digunakan adalah linear kernel karena data sudah cukup terstruktur dan tereduksi dimensinya oleh PCA.

2.5.4. Confusion Matrix

Confusion matrix merupakan metode evaluasi yang digunakan untuk mengukur kinerja suatu model klasifikasi [12]. Matriks ini menampilkan jumlah prediksi yang benar dan salah berdasarkan kondisi aktual dan hasil prediksi dari model. Untuk kasus klasifikasi biner (dua kelas), struktur *confusion matrix* dibagi menjadi empat komponen utama, yaitu *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN). Bentuk umum *confusion matrix* biner dapat dilihat pada tabel berikut [12].

Tabel 1. *Confusion Matrix*

	Prediksi Positif	Prediksi Negatif
Aktual Positif	<i>True positive</i> (TP)	<i>False negative</i> (FN)
Aktual Negatif	<i>False positive</i> (FP)	<i>True negative</i> (TN)

Berdasarkan Tabel 1, dapat dijelaskan bahwa [12]:

1. *True Positive* (TP) adalah jumlah data yang benar-benar positif dan berhasil diprediksi dengan benar sebagai positif.
2. *False Positive* (FP) adalah jumlah data yang sebenarnya negatif, tetapi salah diklasifikasikan sebagai positif.
3. *False Negative* (FN) adalah jumlah data yang sebenarnya positif, namun salah diklasifikasikan sebagai negatif.
4. *True Negative* (TN) adalah jumlah data yang benar-benar negatif dan berhasil diklasifikasikan dengan benar sebagai negatif.

Dari *confusion matrix*, dapat dihitung beberapa matrik evaluasi untuk mengukur kinerja model, seperti akurasi, *precision*, *recall*, dan *F1-score*. Masing-masing rumus dapat dirumuskan sebagai berikut [12]:

1. Akurasi (*Accuracy*) adalah mengukur proporsi prediksi yang benar dari keseluruhan data:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3) [12]$$

Keterangan:

TP = *True positive*

TN = *True negative*

FP = *False positive*

FN = *False negative*

2. Presisi (*Precision*) adalah mengukur ketepatan prediksi positif:

$$Precision = \frac{TP}{TP + FP} \quad (4) [12]$$

Keterangan:

TP = *True positive*

FP = *False positive*

3. *Recall* atau *True Positive Rate* (TPR) adalah mengukur seberapa banyak data positif yang berhasil dikenali:

$$Recall = \frac{TP}{TP + FN} \quad (5) [12]$$

Keterangan:

TP = *True positive*

FN = *False negative*

4. *F1-Score* adalah harmonisasi antara *precision* dan *recall*:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6) [12]$$

Keterangan:

Precision = Nilai *precision* yang dihitung

Recall = Nilai *recall* yang dihitung

Dalam konteks klasifikasi multi-kelas, seperti pada penelitian ini yang melibatkan empat kelas (*healthy*, *rust*, *redrot*, dan *yellow*), *confusion matrix* diperluas ke dalam bentuk matriks persegi berdimensi 4×4. Masing-masing nilai diagonal merepresentasikan jumlah prediksi yang benar untuk setiap kelas, sedangkan nilai *non-diagonal* menunjukkan jumlah kesalahan klasifikasi antar kelas. Perhitungan matrik evaluasi dilakukan dengan pendekatan *one-vs-rest*, yaitu memperlakukan setiap

kelas sebagai positif dan kelas lainnya sebagai negatif secara bergiliran.

Dengan menggunakan *confusion matrix* dan evaluasi matrik turunan lainnya, performa model klasifikasi PCA-SVM dapat diukur secara komprehensif, baik dari segi ketepatan klasifikasi maupun keandalan dalam membedakan antar kelas secara detail.

3. Hasil dan Pembahasan

3.1. Pra-pemrosesan

Dataset terdiri dari 800 citra daun tebu mencakup empat kategori, yaitu *healthy*, *rust*, *redrot*, dan *yellow*. Seluruh citra daun tebu akan diubah ukurannya menjadi 128 x 128 piksel untuk menyamakan dimensi *input*, kemudian dikonversi ke dalam format *grayscale*. Konversi ini bertujuan untuk mendapatkan kualitas visual gambar tinggi [13].

Setiap gambar *grayscale* berukuran 128 x 128 menghasilkan 16.384 piksel fitur awal (128 x 128 = 16.384 per gambar). Setelah melakukan tahap perubahan warna menggunakan *grayscale*, tahap selanjutnya melakukan pelatihan (*training*) dan pengujian (*testing*). Pelatihan dan pengujian dilakukan untuk mempengaruhi nilai akurasi dari suatu pengolahan data [14]. Data pelatihan (*training*) dan data pengujian (*testing*) akan dijadikan 3 bagian perbandingan, yaitu 60:40, 70:30, dan 80:20.

Tabel 2, menyajikan data pelatihan (*training*) dan data pengujian (*testing*) pada masing-masing kategori citra daun tebu.

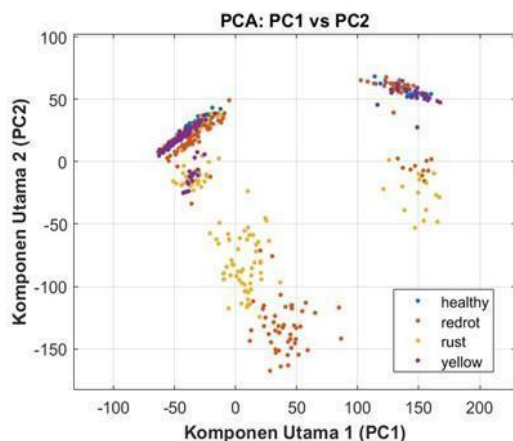
Tabel 2 Kumpulan Pelatihan Dan Pengujian Daun Tebu

Data Pelatihan (%) dan Data Uji (%)	Kategori	Pelatihan	Pengujian
60 : 40	<i>Healthy</i>	120	80
	<i>Rust</i>	120	80
	<i>Redrot</i>	120	80
	<i>Yellow</i>	120	80
	Total	480	320
70 : 30	<i>Healthy</i>	140	60
	<i>Rust</i>	140	60
	<i>Redrot</i>	140	60
	<i>Yellow</i>	140	60
	Total	560	240
80 : 20	<i>Healthy</i>	160	40
	<i>Rust</i>	161	39
	<i>Redrot</i>	160	40
	<i>Yellow</i>	160	40
	Total	641	159

3.2. Reduksi Dimensi dengan PCA

Untuk mengatasi dimensi fitur yang tinggi, digunakan metode *Principal Component Analysis* (PCA) sebagai metode reduksi dimensi yang dapat meminimalisir tingginya ukuran fitur. Hasil reduksi memperoleh jumlah komponen utama yang mempertahankan total variansi sebesar 98% bervariasi tergantung pada proporsi data pelatihan, yaitu sebanyak 296 komponen utama untuk 60%, 339 komponen utama untuk 70%, dan 381 komponen utama untuk 80%.

Visualisasi menggunakan dua komponen utama PCA ditampilkan pada gambar 3 untuk memetakan persebaran data. Berdasarkan visualisasi diberikan, data citra daun tebu (*healthy*, *rust*, *redrot*, dan *yellow*) membentuk suatu kelompok masing-masing, meskipun terdapat sedikit tertimpa antar citra daun tebu tertentu, tetapi mendapatkan hasil persebaran data yang tidak berpenjar dengan kelompoknya sendiri maka visualisasi PCA dapat dinilai cukup baik dalam menangani persebaran data tersebut dengan menggunakan metode *grayscale*.



Gambar 3. Hasil PCA: Komponen 1 dan Komponen 2

3.3. Evaluasi Kinerja Model

Model klasifikasi yang dilakukan menggunakan Support Vector Machine (SVM), dan performanya dievaluasi menggunakan: akurasi, *recall*, *precision*, dan *f1-score*. Pembuatan model tersebut digunakan untuk membandingkan hasil setiap kategori mana yang terbaik dan terendah. Evaluasi dilakukan pada 3 data pelatihan dan pengujian, yaitu 60:40, 70:30, dan 80:20. Berikut tabel hasil evaluasi model.

Tabel 3 Hasil Kinerja Confusion Matrix (Pelatihan 60% dan Pengujian 40%)

Kategori	Akurasi	Precision	Recall	F1-Score
Healthy	73,75%	66,29%	73,75%	69,82%
Redrot	91,25%	96,05%	91,25%	93,59%
Rust	88,75%	92,21%	88,75%	90,45%
Yellow	62,50%	64,10%	62,50%	63,29%
Rata-rata	79,06%	79,66%	79,06%	79,29%

Tabel 4 Hasil Kinerja Confusion Matrix (Pelatihan 70% dan Pengujian 30%)

Kategori	Akurasi	Precision	Recall	F1-Score
Healthy	65,00%	69,64%	65,00%	67,24%
Redrot	95,00%	100,00%	95,00%	97,44%
Rust	93,33%	93,33%	93,33%	93,33%
Yellow	70,00%	62,69%	70,00%	66,14%
Rata-rata	80,83%	81,42%	80,83%	81,04%

Tabel 5 Hasil Kinerja Confusion Matrix (Pelatihan 80% dan Pengujian 20%)

Kategori	Akurasi	Precision	Recall	F1-Score
Healthy	65,00%	70,27%	65,00%	67,53%
Redrot	97,44%	100,00%	97,44%	98,70%
Rust	95,00%	95,00%	95,00%	95,00%
Yellow	70,00%	63,64%	70,00%	66,67%
Rata-rata	81,76%	82,23%	81,86%	81,98%

Berdasarkan tabel hasil kinerja *confusion matrix*, performa klasifikasi oleh masing-masing kategori daun tebu menunjukkan variasi yang cukup signifikan. Hal ini memberikan efektivitas model dalam membedakan setiap kategori daun berdasarkan fitur yang telah diekstraksi menggunakan *grayscale*, PCA, dan diklasifikasikan oleh SVM.

Hasil kinerja *confusion matrix* dari berbagai pelatihan dan pengujian menunjukkan hasil rata-rata yang berbeda setiap adanya perubahan proporsi data pelatihan dan pengujian. Rata-rata akurasi meningkat dari 79,06% (60% pelatihan) menjadi 80,83% (70%), dan 81,76% (80%). Peningkatan terjadi serupa pada *precision* (79,66% → 81,42% → 82,23%), *recall* (79,06% → 80,83% → 81,86%), dan *f1-score* (79,29% → 81,04% → 81,98%). Hasil ini menunjukkan bahwa model mendapatkan performa klasifikasi yang lebih baik dengan peningkatan jumlah data pelatihan.

3.4. Analisis Confusion Matrix

Gambar 4, menunjukkan kategori tertinggi pada *redrot* dengan hasil prediksi benar sebesar 73 data mengklasifikasikan dirinya benar. Namun, kategori rendah menunjukkan pada kategori *yellow* dengan hasil prediksi benar 50. Kesalahan yang terjadi ditunjukkan bahwa kategori *yellow* sulit membedakan dengan kategori *healthy*. Hal ini disebabkan adanya kemiripan fitur akibat konversi *grayscale*.

Confusion Matrix (60% Training)

True Class \ Predicted Class	healthy	redrot	rust	yellow
healthy	59			21
redrot	1	73	4	2
rust	1	3	71	5
yellow	28		2	50

Gambar 4. Confusion matrix Pelatihan 60% dan Pengujian 40%

Gambar 5, menunjukkan hal yang serupa pada hasil kategori tertinggi yaitu *redrot* sebesar 57 data prediksi benar, sementara itu kategori terendah jatuh kepada *healthy* sebesar 39 data memprediksi benar. Kategori *healthy* sulit memprediksi diri sendiri dikarenakan adanya kategori *yellow* yang diprediksi sebesar 21 data.

Confusion Matrix (70% Training)

True Class \ Predicted Class	healthy	redrot	rust	yellow
healthy	39			21
redrot		57	2	1
rust	1		56	3
yellow	16		2	42

Gambar 5. Confusion matrix Pelatihan 70% dan Pengujian 30%

Gambar 6, menunjukkan hasil *confusion matrix* pelatihan 80% dan pengujian 20% yang serupa dengan hasil sebelumnya. Kategori *redrot* tetap menunjukkan hasil yang terbaik dan kategori *healthy* kembali menjadi kategori rendah.

Confusion Matrix (80% Training)

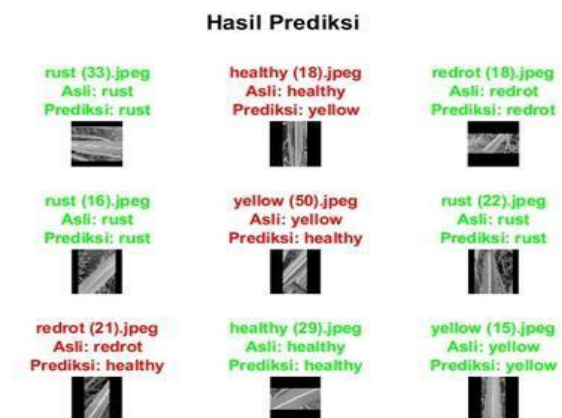
True Class \ Predicted Class	healthy	redrot	rust	yellow
healthy	26			14
redrot		38	1	
rust			38	2
yellow	11		1	28

Gambar 6. Confusion matrix Pelatihan 80% dan Pengujian 20%

Secara keseluruhan dari hasil *confusion matrix* dari gambar 4 hingga gambar 6, mendapatkan kategori *redrot* yang secara konsisten menjadi kategori yang jumlah prediksi benar tertinggi, sedangkan kategori *healthy* dan *yellow* menunjukkan hasil kategori terendah. Hal unik yang terjadi kedua kategori, yaitu *healthy* dan *yellow* sulit diprediksi dikarenakan hilangnya informasi warna pada grayscale.

3.5. Hasil Prediksi

Gambar 7, menunjukkan 9 sampel data yang hasil klasifikasinya dari metode *grayscale*, PCA, dan SVM. Salah satu contoh hasil prediksi menunjukkan bahwa gambar *yellow* diprediksi sebagai *healthy* dan *healthy* diprediksi sebagai *yellow*.



Gambar 7. Hasil Prediksi Daun Tebu

4. Kesimpulan

Berdasarkan hasil eksperimen dilakukan terhadap 800 citra daun tebu dalam empat kategori (*healthy*, *rust*, *redrot*, dan *yellow*) dengan menggunakan metode *grayscale*, *Principal Component Analysis* (PCA), dan *Support Vector Machine* (SVM) dapat disimpulkan beberapa hal:

1. Pra-pemrosesan yang dilakukan berupa *resize* gambar menjadi 128 x 128 piksel dan konversi format gambar ke *grayscale* berhasil diterapkan dan mengurangi kompleksitas dimensi warna.
2. Penerapan PCA sebagai reduksi dimensi yang mampu mempertahankan total variansi sebesar 98,00% citra daun tebu.
3. Visualisasi PCA mendapatkan hasil cukup baik, dimana data saling berkelompok dengan kategorinya masing-masing dan sedikitnya berpisah antar kelompok kategorinya masing-masing.
4. Hasil kinerja *confusion matrix* yang diperlihatkan mendapatkan rata-rata keseluruhan yang terbaik dari setiap kategori citra daun tebu

- sebesar akurasi 81,76%, *precision* 82,23%, *recall* 81,86%, dan *f1-score* 81,98%.
5. *Confusion matrix* menunjukkan akurasi yang bervariasi antara kategori. Kategori *rust* menunjukkan hasil konsisten tertinggi dari berbagai data pelatihan dan data pengujian sedangkan kategori *healthy* dan *yellow* menunjukkan hasil kategori terendah.
 6. Secara keseluruhan, kombinasi model *grayscale*, PCA, dan SVM mendapatkan hasil yang baik pada pelatihan 80% dan pengujian 20% sebesar 81,76%.

REFERENSI

- [1] Nurhajjah, "PENGARUH PEMBERIAN NIPAGIN TERHADAP PERKEMBANGAN PENGGEREK BATANG TEBU BERGARIS (Chilo sacchariphagus)," 2022.
- [2] I. Topan Adib Amrulloh *et al.*, "EVALUASI AUGMENTASI DATA PADA DETEKSI PENYAKIT DAUN TEBU DENGAN YOLOV8," 2024.
- [3] S. Antony Mrefu, Z. Ibna Mannan, and N. Alam, "Enhanced Image-Based Detection and Segmentation of Plant Leaf Diseases Using Grayscale Conversion and Thresholding Techniques," vol. 1, no. 3, pp. 132–139, 2024, doi: 10.69626/cai.2024.0132.
- [4] P. M. Kai, B. M. de Oliveira, and R. M. da Costa, "Deep Learning-Based Method for Classification of Sugarcane Varieties," *Agronomy*, vol. 12, no. 11, Nov. 2022, doi: 10.3390/agronomy12112722.
- [5] X. Liu, B. Zhao, and W. He, "Simultaneous feature selection and classification for data-adaptive Kernel-Penalized SVM," *Mathematics*, vol. 8, no. 10, pp. 1–22, Oct. 2020, doi: 10.3390/math8101846.
- [6] C. N. Govardhini, "Detection of Plant Diseases Using Machine Learning Techniques," 2022. [Online]. Available: www.ijrpr.com
- [7] E. Meirista, M. F. V. Ruslau, Nurhayati, and R. A. Pratama, "Application of Principal Component Analysis (PCA) for Dimensionality Reduction in Watermelon Leaf Classification," *Nusantara Science and Technology Proceedings*, pp. 129–139, May 2025, doi: 10.11594/nstp.2025.4815.
- [8] M. J. Suarez Baron, A. L. Gomez, and J. E. E. Diaz, "Supervised Learning-Based Image Classification for the Detection of Late Blight in Potato Crops," *Applied Sciences (Switzerland)*, vol. 12, no. 18, Sep. 2022, doi: 10.3390/app12189371.
- [9] I. Zeger, S. Grgic, J. Vukovic, and G. Sisul, "Grayscale Image Colorization Methods: Overview and Evaluation," *IEEE Access*, vol. 9, pp. 113326–113346, 2021, doi: 10.1109/ACCESS.2021.3104515.
- [10] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," *Nature Reviews Methods Primers*, vol. 2, no. 1, Dec. 2022, doi: 10.1038/s43586-022-00184-w.
- [11] A. Roy and S. Chakraborty, "Support vector machine in structural reliability analysis: A review," May 01, 2023, *Elsevier Ltd.* doi: 10.1016/j.ress.2023.109126.
- [12] D. Valero-Carreras, J. Alcaraz, and M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Comput Oper Res*, vol. 152, Apr. 2023, doi: 10.1016/j.cor.2022.106131.
- [13] B. S. Hantono *et al.*, "Enhancing Counterfeit Detection with Multi-Features on Secure 2D Grayscale Codes," *Computers*, vol. 12, no. 9, Sep. 2023, doi: 10.3390/computers12090183.
- [14] Baiq Nurul Azmi, Arief Hermawan, and Donny Avianto, "Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4, pp. 281–290, Feb. 2023, doi: 10.35746/jtim.v4i4.298.

Hengky Wisianto, saat ini sebagai Mahasiswa program studi Teknik Informatika Universitas Tarumanagara.

Hizkia Rahuel Anglie, saat ini sebagai Mahasiswa program studi Teknik Informatika Universitas Tarumanagara.

Alfian Effendi, saat ini sebagai Mahasiswa program studi Teknik Informatika Universitas Tarumanagara.