

Identifikasi Kematangan Buah Tomat Menggunakan *Principal Component Analysis* dan *K-Nearest Neighbour* Berdasarkan Citra Warna

Tomato Ripeness Identification Using Principal Component Analysis and K-Nearest Neighbor Based on Color Image

KHAIRANI^{1*}, IMAS SUKAESIH SITANGGANG¹, TOTO HARYANTO¹,
AZIZ KUSTIYO¹

Abstrak

Buah tomat memiliki beberapa tingkat kematangan yang didasarkan pada persentase warna merah pada kulit buah. Tingkat kematangan buah tomat yang telah dipanen berkaitan erat dengan proses distribusinya agar petani memperoleh keuntungan maksimal. Penentuan tingkat kematangan tomat secara manual memiliki kelemahan karena standar yang subjektif dan memakan waktu. Penelitian ini bertujuan untuk mengklasifikasi kematangan tomat berbasis model warna *Hue Saturation Value* (HSV) menggunakan *Principal Component Analysis* (PCA) sebagai ekstraksi ciri dan *K-Nearest Neighbor* (KNN) sebagai pengklasifikasi. Penelitian ini menggunakan 400 citra dengan resolusi spasial 400x400 piksel yang dikelompokkan dalam 5 tingkat kematangan (kelas) yaitu *green*, *turning*, *pink*, *light red* dan *red*. Skenario percobaan dalam penelitian ini adalah mengombinasikan tiga faktor yaitu komponen warna HSV, jumlah *principal component* (PC) dan nilai K pada KNN. Faktor komponen warna HSV terdiri komponen tunggal H, S, dan V serta kombinasinya. Jumlah PC yang dicobakan adalah 5,10,15,...,60 sedangkan nilai K pada KNN ditentukan nilai K = 1, 3, 5, 7, 9 dan 11. Hasil penelitian menunjukkan bahwa komponen warna HV, jumlah PC sebanyak 5 dan K=7 pada KNN menghasilkan nilai akurasi klasifikasi KNN tertinggi sebesar 94%. Hasil penelitian ini meningkat sebesar 1,5% dari penelitian sebelumnya.

Kata Kunci: *hue saturation value*, klasifikasi, *k-nearest neighbor*, *principal component analysis*, tomat.

Abstract

Tomatoes have different levels of ripeness based on the percentage of red color on the fruit's skin. The ripeness level of harvested tomatoes is important for the distribution process to ensure that farmers get maximum profit. Manual determination of tomato ripeness has weaknesses due to subjective standards and being time-consuming. This study aims to classify tomato ripeness based on the *Hue Saturation Value* (HSV) color model using *Principal Component Analysis* (PCA) for feature extraction and *K-Nearest Neighbor* (KNN) as a classifier. The study uses 400 images with a spatial resolution of 400x400 pixels, grouped into 5 ripeness levels: *green*, *turning*, *pink*, *light red*, and *red*. The study explores different factors, including the HSV color component, the number of principal components (PC), and the K value in KNN. The HSV color component factor includes single components H, S, and V, and their combinations. The number of PCs tested ranges from 5 to 60, and the K value in KNN ranges from 1 to 11. The results showed that the HV color component, 5 principal components, and K = 7 in KNN produced the highest KNN classification accuracy value of 94%. This represents a 1.5% increase in accuracy compared to previous studies.

Keywords: *Classification*, *Hue Saturation Value*, *K-Nearest Neighbor*, *Principal Component Analysis*, *Tomato*.

¹ Departemen Ilmu Komputer, FMIPA IPB University;

* Penulis Korespondensi: khairanifahmi88khairani@apps.ipb.ac.id

PENDAHULUAN

Tomat (*Lycopersicon esculentum mill*) adalah jenis komoditas tanaman hortikultura yang banyak digunakan sebagai bahan makanan (Sanjaya *et al.* 2019) dan memiliki pasar yang prospektif (Astrianda 2020). Tomat adalah tanaman multiguna yang memiliki banyak manfaat, yakni sebagai sayuran, bahan masakan, bahan pewarna makanan, bahkan sebagai kosmetik dan obat-obatan (Marliah *et al.* 2012). Tomat sudah menjadi suatu kebutuhan pokok untuk penunjang pangan masyarakat khususnya di Indonesia (Aprilisa dan Sukemi 2019). Produksi buah tomat pada tahun 2022 menurut Badan Pusat Statistik Indonesia adalah 1.116.740 ton, meningkat dari tahun sebelumnya yaitu pada tahun 2021 dengan produksi 1.114.399 ton (BPS 2023).

Buah tomat berasal dari Amerika. Buah ini memiliki beragam bentuk, dimensi, dan pada umumnya berwarna merah ketika buah sudah matang. Salah satu jenis buah tomat yang sering dikonsumsi adalah tomat plum (Salehi *et al.* 2019). Apabila buah tomat dipetik setelah mencapai kematangan (90-100% merah) masa kesegarannya cenderung lebih singkat, serta lebih rentan terhadap kerusakan fisiologis pasca panen dan kerusakan mekanis selama pengangkutan dari kebun ke pasar (Nofrianti 2018). Masa simpan buah tomat pada suhu ruang tanpa kemasan selama 7 hari, sementara untuk penyimpanan dalam lemari pendingin (11-13°C) dengan kertas atau plastik dapat bertahan selama dua minggu. Oleh karena itu buah tomat yang telah dipanen perlu segera didistribusikan agar petani mendapatkan keuntungan maksimal.

Buah tomat memiliki beberapa tingkat kematangan yang didasarkan pada persentase warna merah pada kulit buah. Menurut USDA (1997) tingkat kematangan buah tomat dikelompokkan menjadi 6 level, yaitu *green*, *breakers*, *turning*, *pink*, *light red*, dan *red*. Tingkat kematangan *green* adalah tomat yang berwarna hijau, tingkat kematangan *green* ini dapat berubah dari warna terang ke gelap. Tingkat kematangan *breakers* adalah tomat yang berwarna hijau dengan sedikit warna merah muda pada 10% bagian atasnya. Tingkat kematangan *turning* adalah tomat berwarna hijau terang dan memiliki warna merah muda lebih dari 10% di bagian atasnya. Tingkat kematangan *pink* adalah tomat yang berwarna merah muda pada bagian atasnya antara 30% sampai 60%. Tingkat kematangan *light red* adalah tomat yang berwarna merah muda antara 60% sampai 90% pada atasnya. Tingkat kematangan *red* adalah tomat yang berwarna merah lebih dari 90% sampai warna merah gelap pada atasnya (USDA 1997). Penentuan persentase warna merah pada kulit buah tomat tersebut tidak mudah dilakukan secara visual. Suatu teknik diperlukan agar proses penentuan tingkat kematangan yang konsisten sehingga mengurangi risiko buah tomat menjadi busuk (Riska dan Subekti 2016). Salah satu pendekatan yang dapat dilakukan adalah memanfaatkan pengolahan citra digital dan *machine learning*.

Pengolahan citra digital merupakan bidang ilmu yang berfokus mempelajari cara yang dilakukan dalam pengambilan sebuah informasi tertentu dari sebuah citra. Citra tersebut diolah dan dianalisis sehingga menghasilkan fitur yang informatif (Ilmi *et al.* 2021). Salah satu fitur yang digunakan dalam pengolahan citra digital adalah warna citra. Ruang warna HSV merupakan representasi warna yang terjadi karena adanya kemiripan fitur pada penglihatan manusia (Zhang *et al.* 2018).

Penelitian terkait klasifikasi kematangan buah tomat antara lain dilakukan oleh Astrianda (2020). Penelitian ini melakukan klasifikasi kematangan buah tomat dengan membandingkan beberapa model ruang warna. Penelitian ini menggunakan algoritma SVM dengan akurasi tertinggi berdasarkan model warna CIElab. Penelitian lainnya dilakukan oleh Aprilisa dan Sukemi (2019), yang menggunakan KNN dan model warna RGB. Akurasi tertinggi dari penelitian tersebut adalah 92% dengan input berupa rata-rata dari setiap komponen warna R, G dan B. Riska dan Subekti (2016) melakukan klasifikasi menggunakan model warna gabungan antara RGB dan L*a*b menggunakan algoritma KNN dan Multi-SVM. Akurasi yang dihasilkan dari klasifikasi tersebut adalah 77,79% dengan algoritma KNN dan 77,84% dengan menggunakan SVM. Nilai input yang digunakan pada penelitian tersebut adalah nilai rata-

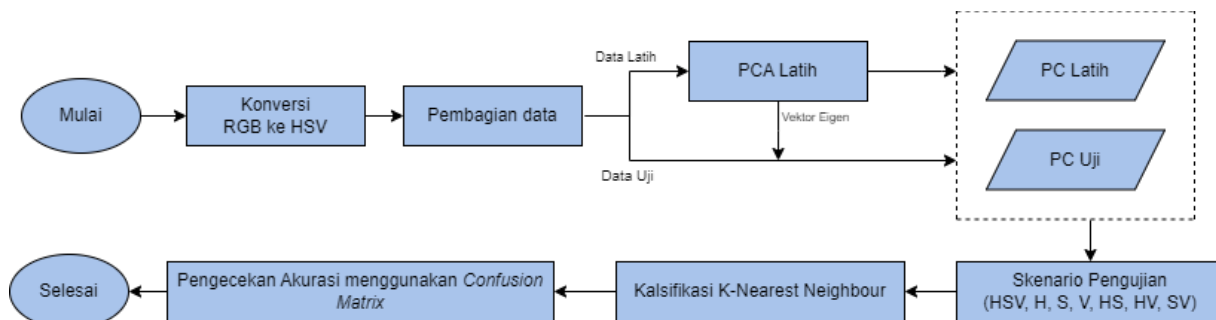
rata dari R,G, dan a*. Sanjaya *et al.* (2019) melakukan penelitian dengan ruang warna HSV. Penelitian ini menggunakan algoritma KNN dengan input rata-rata dari H,S, dan V dengan akurasi 92,5%.

Berdasarkan penelitian-penelitian tersebut, input yang digunakan adalah berupa rata-rata nilai dari setiap komponen ruang warna RGB atau HSV. Nilai rata-rata sangat rentan terhadap nilai pencilan. Rata-rata bisa menjadi terlalu kecil atau menjadi terlalu besar tergantung input dari nilai ruang warna yang digunakan. Oleh karena itu dibutuhkan sebuah metode ekstraksi fitur yang lebih baik dibandingkan dengan nilai rata-rata. Salah satu metode yang dapat digunakan adalah *principal component analysis* (PCA). Tujuan PCA adalah menangkap variasi total pada data citra dan merepresentasikan ke dalam variabel yang jumlahnya lebih sedikit (Suryadi 2015). Dengan demikian, hasil ekstraksi fitur PCA lebih mampu mewakili keseluruhan data dibandingkan dengan menggunakan nilai rata-rata. Model warna yang digunakan adalah HSV karena berdasarkan penelitian Sanjaya *et al.* (2019) mampu menghasilkan akurasi 92,5%. Adapun algoritma yang digunakan untuk klasifikasi adalah KNN seperti pada penelitian-penelitian sebelumnya.

Penelitian ini akan mengidentifikasi citra buah tomat menggunakan KNN dengan praproses PCA berdasarkan model warna HSV. Penggunaan PCA sebagai praproses, diharapkan dapat meningkatkan akurasi klasifikasi algoritma KNN.

METODE

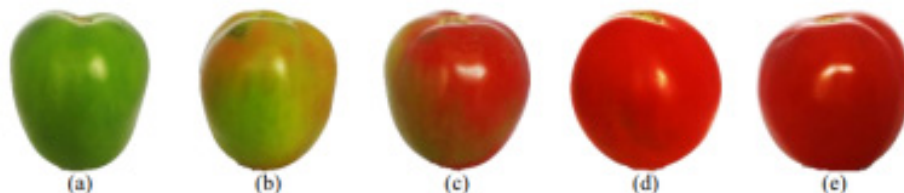
Tahapan pada penelitian ini disajikan pada Gambar 1. Secara umum tahapan penelitian terdiri dari konversi citra RGB ke HSV, praproses PCA, klasifikasi KNN dan evaluasi hasil klasifikasi.



Gambar 1 Tahapan penelitian

Data Citra

Penelitian ini menggunakan buah tomat jenis *plum* (*Lycopersicon lycopersicum*) dengan lima kategori tingkat kematangan buah tomat yaitu *green*, *turning*, *pink*, *light red* dan *red* seperti pada Gambar 2.



Gambar 2 Tingkat kematangan buah tomat (a) green, (b) turning, (c) pink, (d) light red, (e) red (Sanjaya *et al.* 2019)

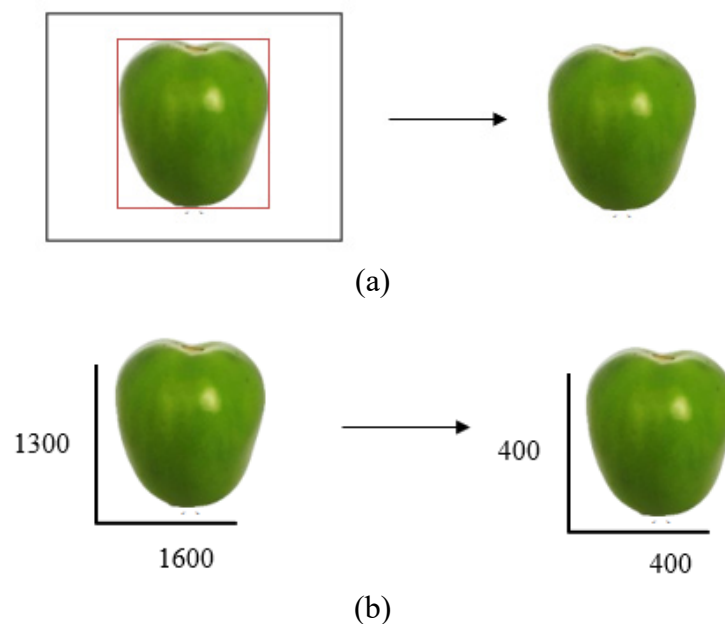
Sejumlah 400 citra diakuisisi pada studi ini. Untuk setiap kelas, terdapat 20 buah tomat. Terdapat 80 sampel citra pada total buah tomat yang digunakan sebagai disajikan pada Tabel 1. Pada penelitian ini dilakukan pengambilan gambar pada 4 sisi buah tomat, sehingga diperoleh 4 data citra untuk setiap buah tomat.

Tabel 1 Jumlah citra buah tomat

No	Nama Atribut	Jumlah buah tomat	Jumlah citra buah tomat
1	Green	20 Buah	80 citra
2	Turning	20 Buah	80 citra
3	Pink	20 Buah	80 citra
4	Light red	20 Buah	80 citra
5	Red	20 Buah	80 citra
Jumlah		100 Buah	400 citra

Praproses Data

Tahap ini dilakukan untuk menghilangkan *noise* pada citra. Proses yang dilakukan pada proses ini adalah menghilangkan *background* dan kemudian melakukan *cropping* yang berfokus pada buah tomat. Citra hasil *cropping* berukuran 1300×1600 piksel. Proses selanjutnya adalah melakukan *resize* sesuai dengan yang telah ditentukan yaitu 400×400 . Tahapan praproses ini disajikan pada Gambar 3.

Gambar 3 (a) Proses *cropping* buah tomat, (b) Proses *resize* citra buah tomat

Konversi Citra RGB ke HSV

Penelitian ini menggunakan ekstraksi fitur ruang warna HSV. HSV terdiri dari tiga komponen yaitu *Hue* (H) yang mewakili warna, *Saturation* (S) mewakili tingkat kepekatan warna, dan *Value* (V) yang mewakili tingkat kecerahan (Indriani *et al.* 2017). Penelitian diawali dengan pengambilan nilai RGB dari citra yang telah melalui praproses data. Nilai RGB merepresentasikan nilai warna merah, hijau, dan biru dengan rentang nilai antara 0 sampai 255 (Pratama *et al.* 2019). Fitur HSV diperoleh dari konversi citra RGB tersebut. Setelah dilakukan proses konversi, selanjutnya fitur HSV dinormalisasi. Hal ini dilakukan untuk mengubah nilai HSV ke rentang 0 sampai 1 agar lebih mudah dalam proses komputasi.

Model HSV adalah turunan dari RGB, tetapi HSV lebih unggul dibandingkan RGB. HSV bisa mengekspresikan bayangan warna, kontras warna, derajat warna dan rona warna. HSV memiliki tiga komponen utama yaitu:

1. *Hue* mewakili warna dasar atau warna sebenarnya yang memiliki rentang dari 0° hingga 360° . Titik 0° merupakan warna yang bervariasi dari merah, kuning, cyan, biru dan magenta kemudian kembali lagi ke warna merah (Hasym dan Susilawati 2021). Perhitungan nilai *Hue* dilakukan dengan Persamaan 1.
2. *Saturation* mewakili tingkat kemurnian pada satu warna yang memiliki rentang 0 hingga

1. Nilai 0 merupakan warna yang bernuansa abu-abu hingga nilai 1 yang merupakan komponen putih (Irawan dan Rachmawanto 2022). Perhitungan nilai *Saturation* dilakukan dengan Persamaan 2.
3. *Value* atau disebut kecerahan dari setiap warna yang mewakili seberapa gelap atau seberapa cerah warna tersebut. *Value* memiliki rentang nilai 0% hingga 100%. Nilai 0% mewakili warna hitam, ini berarti bahwa semakin tinggi nilai *Value* maka semakin cerah (Anggriawan *et al.* 2017). Perhitungan nilai *Value* dilakukan dengan Persamaan 3.

$$H = \arctan \left\{ \frac{\sqrt{3}(G-B)}{(R-G)+(R-B)} \right\} \quad (1)$$

$$S = 1 - \frac{\min(R,G,B)}{v} \quad (2)$$

$$V = \frac{R+G+B}{3} \quad (3)$$

Principal Component Analysis (PCA)

Setelah dilakukan proses ekstraksi fitur warna menggunakan model warna HSV, selanjutnya dilakukan proses ekstraksi ciri menggunakan PCA. Pada penelitian ini persentase data latih yang digunakan adalah sebesar 80% dan persentase data uji sebesar 20%. Hasil yang diperoleh pada tahapan ini adalah nilai *principal component* (PC) atau vektor perwakilan citra. Nilai vektor yang didapat akan menjadi ciri pada suatu citra dalam penelitian klasifikasi buah tomat plum.

PCA membutuhkan input data dengan atribut *zero-mean* pada setiap fitur. *Zero-mean* pada setiap fitur dapat diperoleh dengan mengurangi semua nilai data dengan rata-rata. Ilustrasi PCA disajikan dengan *data set X* dengan dimensi .

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ x_{M1} & x_{M1} & x_{11} & x_{MN} \end{bmatrix}$$

Untuk karakteristik ke-*j*, semua nilai kolom dikurangi dengan rata-rata, dihitung dengan Persamaan 4. Kemudian dilakukan proses untuk mendapatkan matriks *covariance* dari matriks X' , yaitu C , dengan Persamaan 5 dengan $X'=[x'_i]$, $i = 1,2,\dots, M, j = 1,2,\dots, N$, dan \bar{x}_j nilai rata-rata kolom ke-*j*.

$$x'_{ij} = x_{ij} - \bar{x}_j \quad (4)$$

$$C = \frac{1}{M} X' \cdot X'^T \quad (5)$$

$$C - \lambda I = 0 \quad (6)$$

$$(C - \lambda I)v = 0 \quad (7)$$

Dengan C adalah matriks *covariance*, I adalah matrik identitas, λ adalah nilai *eigen* dan v adalah *vector eigen*. Nilai *eigen* dan vektor *eigen* dari matriks *covariance* dihitung dengan menggunakan Persamaan 6 dan 7. Nilai *eigen* terbesar yang cocok dengan nilai vektor *eigen* terbesar dipilih sebagai *principal component*, vektor *eigen* yang urutkan dari yang terbesar

ke yang terkecil dipilih menjadi vektor fitur Persamaan 8 dan *principal component* dengan X' sebagai rata-rata dicari dengan Persamaan 9. Selanjutnya langkah yang dilakukan adalah transformasi data untuk menghasilkan data PCA dengan X sebagai data sesuai Persamaan 10.

$$\text{Vector Fitur} = (eig_1, eig_2, eig_3 \dots eig_n) \quad (8)$$

$$PC = X' \times v \quad (9)$$

$$PCA \text{ data} = PC^T \times X'^T \quad (10)$$

K-Nearest Neighbor (K-NN)

Algoritma *K-Nearest Neighbor* (KNN) merupakan algoritma yang bersifat *non-parametric* dan *lazy learning* (Ningsih and Cholidha 2022). Algoritma ini digunakan untuk menyelesaikan masalah klasifikasi objek (Rivki dan Bachtiar 2017). Algoritma KNN merupakan teknik klasifikasi data baru berdasarkan K tetangga terdekat yang bermakna bahwa data tersebut memiliki kesamaan (Adenugraha *et al.* 2022). Kedekatan tersebut dilihat dari jaraknya yang menggunakan jarak *euclidean* (Persamaan 11).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (11)$$

dengan :

$d(x, y)$ = Jarak *euclidean* antara *vector* x dan y

i = Indeks fitur

n = Jumlah fitur dalam *vector* x dan y

Input untuk algoritma KNN merupakan kombinasi dari komponen warna dari hasil PCA yang menggunakan skenario:

1. Satu Komponen warna : H, S, dan V
2. Dua komponen warna : HS, HV dan SV
3. Tiga komponen warna : HSV

Skenario Percobaan

Penelitian ini melibatkan 3 faktor yaitu kombinasi komponen-komponen model warna HSV, jumlah *principal component* (PC) sebagai input KNN, dan nilai K pada KNN dengan rincian sebagai berikut:

- a. Kombinasi komponen warna HSV ada 7 terdiri dari H, S, V, HS, HV, SV dan HSV
- b. Jumlah PC ada 12 yaitu 5, 10, 15, ..., 60
- c. Nilai K pada KNN ada 6 yaitu $K = 1, 3, 5, 7, 9, 11$

Dengan demikian terdapat sebanyak $7 \times 12 \times 6$ percobaan atau total terdapat 504 percobaan.

Evaluasi

Evaluasi dilakukan berdasarkan skenario percobaan yang telah disusun. Peran dari setiap faktor yaitu kombinasi komponen-komponen warna HSV, jumlah PC dan nilai K pada KNN akan dievaluasi terhadap akurasi klasifikasi KNN. Dengan demikian dapat diketahui kombinasi dari ketiga faktor yang menghasilkan akurasi klasifikasi KNN yang tertinggi. Untuk memudahkan evaluasi, digunakan *confusion matrix multiclass* (Tharwat 2021) karena dalam penelitian ini terdapat 5 level tingkat kematangan buah tomat atau 5 kelas. *Multiclass* pada penelitian ini adalah kelas *green*, *turning*, *pink*, *light red*, dan *red* seperti yang ditunjukkan pada Tabel 2. TP_A merupakan *True positive* kelas A yang merupakan jumlah sampel kelas A yang terverifikasi benar pada kelas A. FA_B merupakan merupakan sampel yang terverifikasi salah (sampel kelas

A yang terverifikasi ke kelas B). *False negative* untuk kelas A adalah jumlah sampel kelas A yang salah diverifikasi sebagai kelas B, C, D dan E ($F_{NA} = F_{AB} + F_{AC} + F_{AD} + F_{AE}$). Untuk menghitung akurasi digunakan Persamaan 12 (Tharwat 2021).

Setelah diperoleh kombinasi faktor yang menghasilkan akurasi KNN tertinggi, selanjutnya akan dianalisis kesalahan klasifikasi yang terjadi. Analisis tersebut didasarkan pada *confusion matrix multiclass* (Tharwat 2021) berupa *false negative* untuk setiap kelas ($F_{NA}, F_{NB}, F_{NC}, F_{ND}$ dan F_{NE})

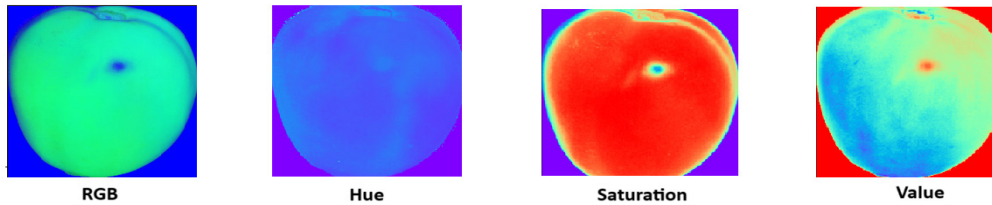
$$Accuracy = \frac{1}{N}(TP_A + TP_B + TP_C + TP_D + TP_E) \tag{12}$$

Tabel 2 *Confusion matrix*

<i>Predicted Class</i>	<i>Actual Class</i>				
	<i>Green</i>	<i>Turning</i>	<i>Pink</i>	<i>Light Red</i>	<i>Red</i>
<i>Green</i>	TP_A	F_{BA}	F_{CA}	F_{DA}	F_{EA}
<i>Turning</i>	F_{AB}	TP_B	F_{CB}	F_{DB}	F_{EB}
<i>Pink</i>	F_{AC}	F_{BC}	TP_C	F_{DC}	F_{EC}
<i>Light Red</i>	F_{AD}	F_{BD}	F_{CD}	TP_D	F_{ED}
<i>Red</i>	F_{AE}	F_{BE}	F_{CE}	F_{DE}	TP_E

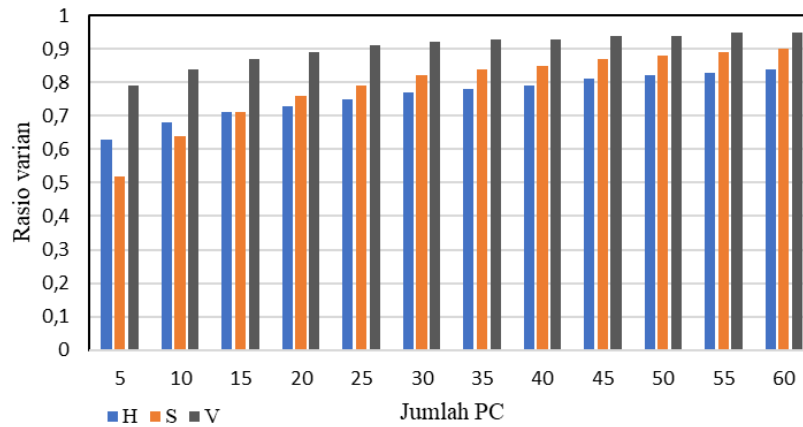
HASIL DAN PEMBAHASAN

Konversi model warna RGB ke model warna HSV dilakukan pada penelitian ini. RGB yang hanya berpusat pada warna memberikan akurasi yang rendah karena dipengaruhi oleh cahaya, kualitas gambar dan bayangan. Oleh karena itu, diperlukan HSV untuk memberikan penilaian yang lebih baik. HSV terdiri dari *hue* sebagai perwakilan warna, *saturation* sebagai perwakilan cahaya dan *value* sebagai kepekatan warna. Pada Gambar 4 dapat dilihat bahwa HSV memberikan gambaran citra yang lebih beragam dari pada RGB, sehingga hasil yang diberikan lebih baik.



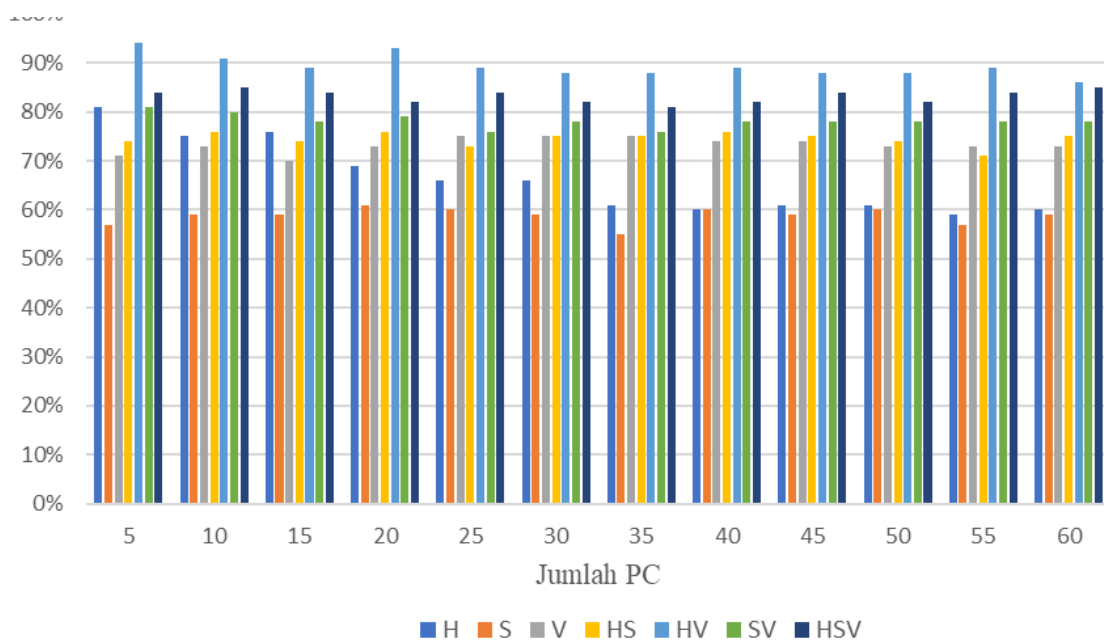
Gambar 4 Contoh citra warna HSV beserta komponennya (a) HSV (b) Hue (c) Saturation (d) Value

Ukuran citra ketika diterapkan PCA adalah 400x400 piksel. Teknik PCA diterapkan pada setiap komponen HSV sehingga terdapat *principal component* (PC) untuk komponen H, PC untuk komponen S dan PC untuk komponen V. Jumlah PC yang diamati untuk setiap komponen tersebut adalah 5,10,15,...,60. Pada Gambar 5 disajikan rasio varian (kontribusi) dari setiap jumlah PC dari setiap komponen HSV. Komponen warna V memiliki rasio varian tertinggi, sedangkan komponen H memiliki rasio varian terendah. Dengan jumlah PC sebesar 20, komponen warna V memiliki rasio varian mendekati 0,9 (90%) sedangkan kedua komponen warna yang lain memiliki rasio varian lebih dari 0,7 (70%).



Gambar 5 Rasio varian jumlah PC untuk setiap komponen HSV

Pada Gambar 6 disajikan akurasi KNN dengan $K = 7$ pada berbagai jumlah PC untuk semua kombinasi komponen warna HSV. Untuk komponen warna tunggal H, S dan V, komponen H menghasilkan akurasi tertinggi sedangkan komponen warna S memiliki akurasi terendah untuk seluruh jumlah PC yang dicobakan. Untuk seluruh kombinasi komponen warna, kombinasi HV menghasilkan akurasi tertinggi untuk semua jumlah PC. Selain itu, dari Gambar 6 juga dapat dilihat bahwa jumlah PC lebih dari 20 tidak menambah akurasi KNN. Untuk kombinasi warna HV dengan jumlah PC 20 berarti bahwa jumlah input untuk KNN adalah sebanyak 40 (masing-masing 20 untuk komponen H dan komponen V). Akurasi KNN pada semua kombinasi komponen warna HSV dengan jumlah PC maksimal 20 disajikan pada Tabel 3. Dari Tabel 3 dapat dilihat bahwa akurasi KNN tertinggi diperoleh ketika jumlah PC sebesar 5 dengan kombinasi komponen warna HV.

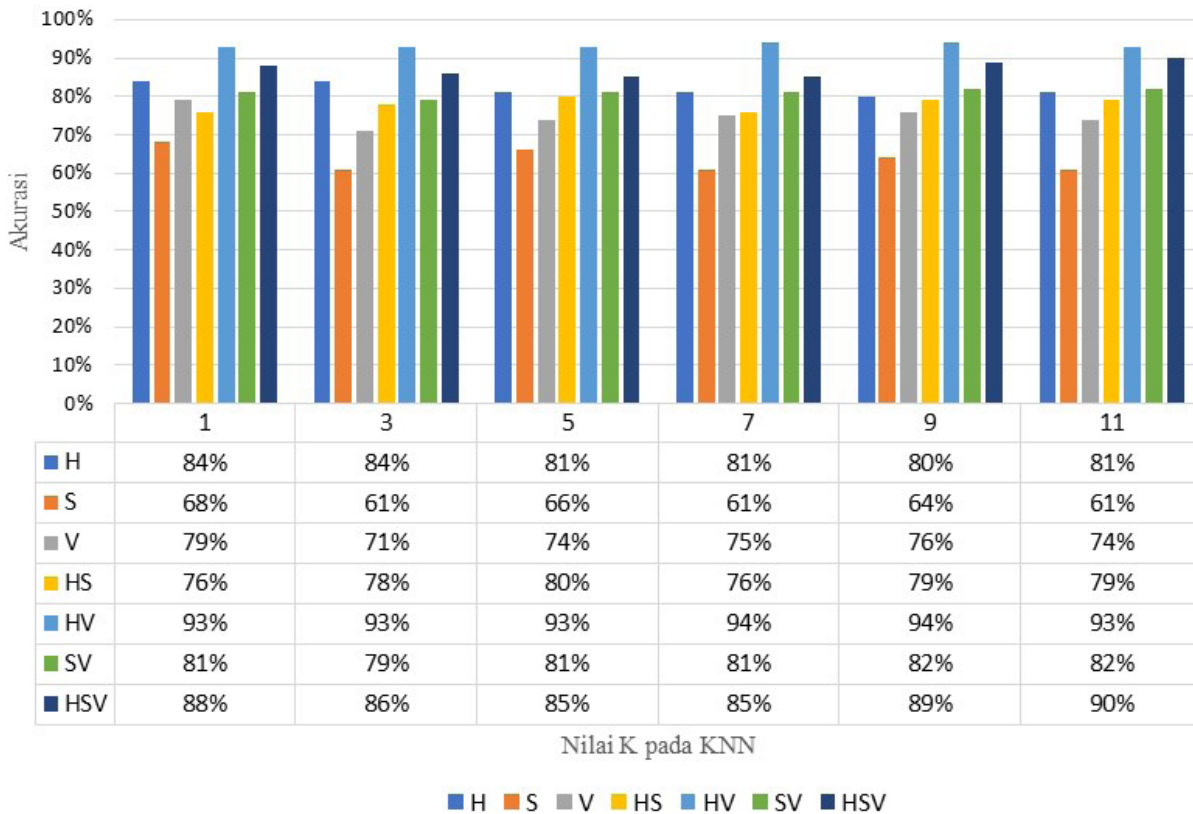


Gambar 6 Akurasi KNN untuk $K = 7$ pada berbagai jumlah PC dan kombinasi komponen warna HSV

Tabel 3 Akurasi KNN untuk $K = 7$ dengan jumlah PC maksimum 20 dan kombinasi komponen warna HSV

Nilai PC	H	S	V	HS	HV	SV	HSV
5	81%	57%	71%	74%	94%	81%	84%
10	75%	59%	72%	76%	91%	80%	85%
15	76%	59%	70%	74%	89%	78%	84%
20	69%	61%	72%	76%	93%	79%	82%

Berbeda dengan Gambar 6 yang hanya menyajikan akurasi KNN untuk $K = 7$ untuk semua kombinasi komponen warna dan seluruh jumlah PC, pada Gambar 7 disajikan akurasi tertinggi KNN pada semua nilai K . Penambahan nilai K pada KNN, untuk komponen warna tunggal (H, S dan V) cenderung menurunkan akurasi KNN, sedangkan untuk kombinasi komponen dua warna cenderung meningkatkan akurasi KNN. Penambahan nilai K untuk kombinasi tiga warna (HSV) pada awalnya menurunkan akurasi KNN ($K = 1, 3$ dan 5) kemudian meningkatkan akurasi KNN ($K = 7, 9$ dan 11). Akurasi KNN tertinggi diperoleh pada saat $K = 7$ dan $K = 9$ dengan input KNN berupa kombinasi HV. Berdasarkan data yang ada, akurasi KNN sebesar 94% untuk $K = 9$ terjadi ketika jumlah PC sebanyak 5. Dengan demikian, bila hasil ini digabungkan dengan hasil pada Tabel 3, jumlah PC yang menghasilkan akurasi KNN tertinggi adalah sebanyak 5 PC.



Gambar 7 Akurasi KNN terbaik dari pengujian seluruh skenario

Selanjutnya untuk mengetahui kesalahan prediksi KNN dengan akurasi tertinggi sebesar 94% pada $K = 7$, disajikan *confusion matrix* pada Tabel 4. Dari 80 data yang diujikan terdapat 5 data kelas *pink* yang salah diklasifikasikan sedangkan kelas-kelas yang lain diklasifikasikan dengan benar. Kesalahan klasifikasi KNN yang terjadi adalah kelas *pink* diklasifikasikan sebagai kelas *green*, *turning* dan *red*. Kesalahan klasifikasi KNN tersebut kemungkinan disebabkan adanya kesamaan fitur warna HV dari kelas *pink* dengan ketiga kelas tersebut karena akuisisi citra dari empat sisi yang berbeda. Tomat plum kelas *pink* masih memungkinkan memiliki komponen warna yang sama dengan ketiga kelas tersebut pada sisi yang berbeda.

Tabel 4 *Confusion matrix* dengan nilai akurasi 94%

Predicted Class	Actual Class				
	Green	Turning	Pink	Light Red	Red
Green	16	0	1	0	0
Turning	0	16	2	0	0
Pink	0	0	11	0	0
Light Red	0	0	0	16	0
Red	0	0	2	0	16

Pada penelitian sebelumnya dengan *data set* yang sama, penelitian dilakukan dengan input KNN berupa rata-rata komponen H, rata-rata komponen S dan rata-rata komponen V (Sanjaya *et al.* 2019). Penelitian tersebut menerapkan variasi ukuran citra dari 100x100 sampai dengan 1000x1000 piksel. Akurasi tertinggi dari KNN pada penelitian tersebut sebesar 92,50% dengan ukuran citra 1000x1000 piksel dan nilai K pada KNN sebesar 3 (Sanjaya *et al.* 2019). Pada penelitian ini, akurasi tertinggi KNN adalah sebesar 94%, meningkat sebanyak 1,50% dibandingkan dengan penelitian Sanjaya *et al.* (2019). Peningkatan ini kemungkinan disebabkan oleh PCA sebagai teknik ekstraksi ciri serta penambahan fitur berupa kombinasi komponen warna HV. Untuk ukuran citra, pada penelitian ini hanya digunakan satu ukuran citra yaitu 400x400 piksel dan tidak dilakukan percobaan dengan berbagai ukuran citra seperti hanya penelitian Sanjaya *et al.* (2019).

SIMPULAN

Hasil penelitian menunjukkan bahwa komponen *Hue* dalam model warna HSV memiliki peran paling penting dalam menentukan akurasi KNN dalam identifikasi tingkat kematangan buah tomat plum. Akurasi KNN mencapai nilai tertinggi yaitu 94% ketika input dari KNN merupakan kombinasi komponen *Hue* dan *Value*.

Penggunaan teknik PCA sebagai teknik ekstraksi ciri dari komponen warna HSV mampu meningkatkan akurasi KNN sebesar 1,5% dibandingkan hasil penelitian tanpa menggunakan teknik PCA. Hasil penelitian menunjukkan bahwa 5 *principal component* per komponen warna HSV menghasilkan akurasi klasifikasi KNN yang tertinggi.

DAFTAR PUSTAKA

- Adenugraha SP, Arinal V, Mulyana DI. 2022. Klasifikasi Kematangan Buah Pisang Ambon Menggunakan Metode KNN dan PCA Berdasarkan Citra RGB dan HSV. *J Media Inform Budidarma*. 6(1):9. doi:10.30865/mib.v6i1.3287.
- Anggriawan MA, Ichwan M, Utami DB. 2017. Pengenalan Tingkat Kematangan Tomat Berdasarkan Citra Warna Pada Studi Kasus Pembangunan Sistem Pemilihan Otomatis. 3:550–564. doi:https://doi.org/10.28932/jutisi.v3i3.695.
- Aprilisa S, Sukemi. 2019. Klasifikasi Tingkat Kematangan Buah Tomat Berdasarkan Fitur Warna. *Pros Annu Res Semin 2019*. 5(1):978–979. https://repository.unsri.ac.id/60988/1/Proseeding-ARS.pdf.
- Astrianda N. 2020. Klasifikasi Kematangan Buah Tomat Dengan Variasi Model Warna Menggunakan Support Vector Machine. *Vocat Educ technolgy J*. 1(2):44–51. doi:https://doi.org/10.38038/vocatech.v1i2.27.
- BPS. 2023. Statistik Indonesia Catalog : 1101001. Volume ke-1101001. https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html.
- Hasym IE, Susilawati I. 2021. Klasifikasi Jenis Ikan Cupang Menggunakan Algoritma Principal Component Analysis (PCA) Dan K-Nearest Neighbors (KNN). *KONSTELASI Konvergensi Teknol dan Sist Inf*. 1(1):168–179. doi:10.24002/konstelasi.v1i1.4242.
- Ilmi A, Razka MH, Wiratomo DS, Prasvita DS. 2021. Klasifikasi Tingkat Kematangan Buah Apel Berdasarkan Fitur Warna Menggunakan Algoritma K-Nearest Neighbor dan Ekstraksi Warna HSV. 2 September:176–182. https://conference.upnvj.ac.id/index.php/senamika/article/view/1633.
- Indriani OR, Kusuma EJ, Sari CA, Rachmawanto EH, Setiadi DRIM. 2017. Tomatoes classification using K-NN based on GLCM and HSV color space. *Proc - 2017 Int Conf Innov Creat Inf Technol Comput Intell IoT, ICITech 2017*. 2018-Janua:1–6. doi:10.1109/INNOCIT.2017.8319133.
- Irawan C, Rachmawanto EH. 2022. Ekstraksi Hsv Dan Glcm Dalam Metode K-Nn Untuk Klasifikasi Tingkat Kematangan Buah Mengkudu. *Pros Snast*. November:D16-25. doi:10.34151/prosidingsnast.v8i1.4150.

- Marliah A, Hayati M, Muliansyah I. 2012. Pemanfaatan Pupuk Organik Cair Terhadap Pertumbuhan Dan Hasil Beberapa Varietas Tomat (*Lycopersicum esculentum L .*). *Agrista*. 16(3):122–128. <https://jurnal.usk.ac.id/agrista/article/view/656/564>.
- Ningsih L, Cholidhazia P. 2022. Classification Of Tomato Maturity Levels Based on RGB And HSV Colors Using KNN Algorithm. *RIGGS J Artif Intell Digit Bus*. 1(1):25–30. doi:10.31004/riggs.v1i1.10.
- Nofrianti D. 2018. *Penanganan Pascapanen Tomat*. Izhar L, Primilestari S, editor. Jambi: Balai Pengkajian Teknologi Pertanian Jambi Badan Penelitian Dan Pengembangan Pertanian Kementerian Pertanian. <https://repository.pertanian.go.id/bitstreams/555bcdbc-efd8-4eac-bec3-fa155b4b570c/download>.
- Pratama R, Assagaf AF, Tempola F. 2019. Deteksi Kematangan Buah Tomat Berdasarkan Fitur Warna Menggunakan Metode Transformasi Ruang Warna HIS. *JIKO*. 2(2):81–86. doi:<https://doi.org/10.33387/jiko.v2i2.1318>.
- Riska SY, Subekti P. 2016. Klasifikasi Level Kematangan Buah Tomat Berdasarkan Fitur Warna Menggunakan Multi-SVM. *J Ilm Inform*. 1(1):39–45. doi:<https://doi.org/10.35316/jimi.v1i1.442>.
- Rivki M, Bachtiar AM. 2017. Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *J Sist Inf*. 13(1):31. doi:10.21609/jsi.v13i1.500.
- Salehi B, Sharifi-Rad R, Sharopov F, Namiesnik J, Roointan A, Kamle M, Kumar P, Martins N, Sharifi-Rad J. 2019. Beneficial effects and potential risks of tomato consumption for human health: An overview. *Nutrition*. 62:201–208. doi:10.1016/j.nut.2019.01.012.
- Sanjaya S, Lisa.PM, Kurnia.GS, Yanto F, Syafria F. 2019. K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue , Saturation , and Value Colors. 2 September 2022:101–106. doi:10.24014/ijaidm.v2i2.7975.
- Suryadi A. 2015. Sistem Pengenalan Wajah Menggunakan Metode Principal Component Analysis (PCA) Dengan Algoritma Fuzzy C-Means (FCM). *J Pendidik Mat*. 4(2):58–65. http://e-mosharafa.org/index.php/mosharafa/article/view/mv4n2_2/194.
- Tharwat A. 2021. Classification assessment methods. 17(1):168–192. doi:10.1016/j.aci.2018.08.003.
- USDA. 1997. United States Standards for Grades of Fresh Tomatoes. *USDA*. 1991 January:1–13. https://www.ams.usda.gov/sites/default/files/media/Tomato_Visual_Aids%5B1%5D.pdf.
- Zhang T, Hu HM, Li B. 2018. A Naturalness Preserved Fast Dehazing Algorithm Using HSV Color Space. *IEEE Access*. 6:10644–10649. doi:10.1109/ACCESS.2018.2806372.