

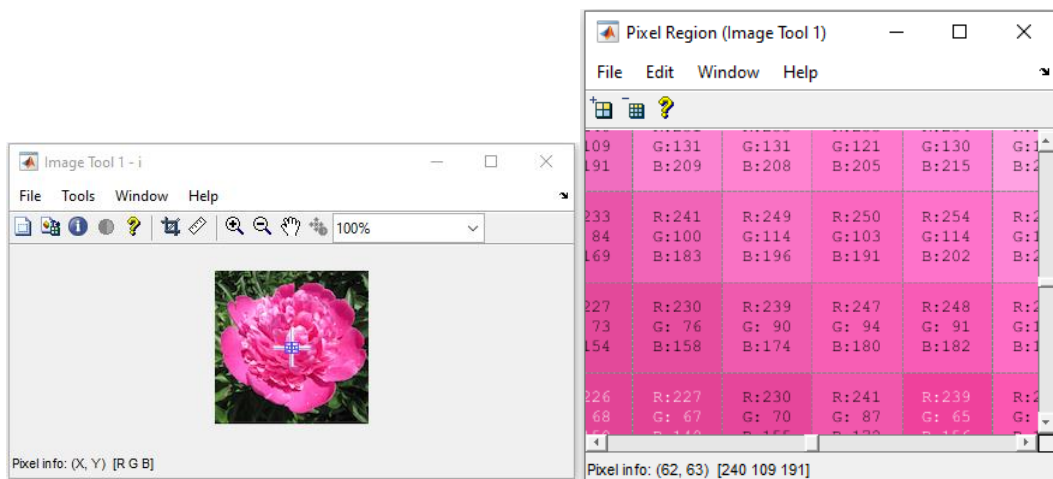
BAB II

TINJAUAN PUSTAKA

2.1. Citra RGB

Citra RGB merupakan citra yang terdiri dari 3 warna yaitu, *red* (merah), *green* (hijau), dan *blue* (biru). Setiap piksel yang ada pada citra RGB terdiri dari 3 kombinasi warna tersebut dengan masing-masing nilai intensitasnya berada diantara 0-255. Setiap informasi piksel RGB disimpan ke dalam 1 *byte* data, jadi 8 bit pertama menyimpan nilai merah, yang kedua hijau, dan yang ketiga biru, sehingga dalam 1 piksel citra RGB mempunyai ukuran 24 bit.

Banyaknya kombinasi warna yang mungkin dihasilkan adalah $256 \times 256 \times 256 = 16.777.216$ warna. Berdasarkan Gambar 2.1 gambaran dari citra RGB dan matriks penyusunnya adalah:



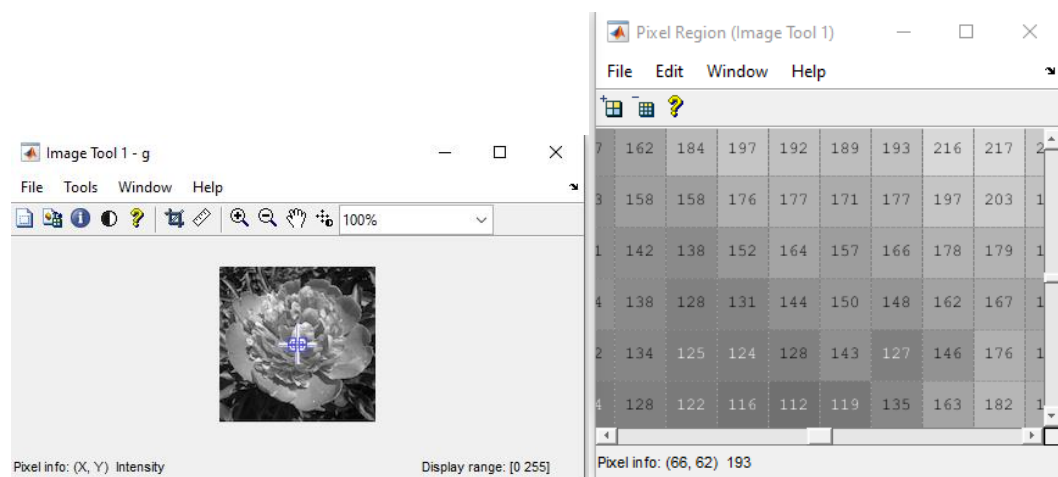
Gambar 2.1 Citra RGB dan Matriks Penyusunnya

2.2. Citra *Grayscale*

Citra *grayscale* merupakan citra yang mempunyai intensitas piksel yang menunjukkan nilai derajat keabuan. Nilai piksel yang dimiliki adalah 0-255 yang ada pada 1 kanal warna. Nilai piksel 0 menunjukkan warna hitam dan nilai piksel 255 menunjukkan warna putih. Jadi, warna abu-abu yang dihasilkan berada diantara *range* nilai tersebut. Cara yang digunakan untuk menghitung konversi dari citra RGB menjadi citra *grayscale* dengan hitungan matematis adalah:

$$\text{Grayscale} = (0.2989 \times \text{red}) + (0.5870 \times \text{green}) + (0.1140 \times \text{blue})$$

Berdasarkan Gambar 2.2 gambaran dari citra *grayscale* dan matriks penyusunnya adalah:



Gambar 2.2 Citra *Grayscale* dan Matriks Penyusunnya

2.3. *Principal Component Analysis*

Beberapa Principal Component (PC) sudah cukup untuk menjelaskan struktur data asli. Jika data dimensi asli sulit untuk direpresentasikan melalui grafik, maka dengan 2 PC atau 1 PC bisa dicitrakan melalui grafik (Santosa 2007). Jika

terdapat contoh data dengan susunan seperti pada Gambar 2.3, dengan n adalah jumlah variabel/atribut dan m adalah banyaknya observasi.

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \dots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix}$$

Gambar 2.3 Contoh Susunan Data PCA

Kemudian data tersebut dapat ditransformasikan ke dalam suatu kolom, misalkan dengan cara centering yaitu mengurangi setiap data dengan rata-rata dari setiap atribut yang ada, dengan rumus seperti pada Gambar 2.4 di bawah ini:

$$\hat{X} = X - \bar{X}$$

Gambar 2.4 Rumus *centering*

Dapat dijelaskan bahwa \hat{X} adalah vektor hasil setelah centering, X adalah vektor kolom dan \bar{X} merupakan rata-rata dari kolom yang bersangkutan. Proses tersebut dilakukan untuk semua kolom dari $I = 1$ sampai $I = n$, dan untuk mendapatkan besaran baru digunakan matriks *Covariance* (C) seperti pada Gambar 2.5.

$$C = \frac{\hat{X}^T \hat{X}}{m - 1}$$

Gambar 2.5 Matriks *Covariance*

Di dalam *pattern recognition*, fitur-fitur suatu citra menggunakan PCA direpresentasikan sebagai *eigenvectors*. *Eigenvectors* merupakan kumpulan hubungan karakteristik-karakteristik dari suatu citra untuk mengenali citra tersebut secara spesifik. Setelah matrik *covariance* dihitung, langkah berikutnya adalah mencari *eigenvalue* dan *eigenvector*. *Eigenvalue* yang didapat diurutkan mulai yang terbesar sampai dengan yang terkecil, serta *eigenvector* yang bersesuaian dengan *eigenvalue* tersebut juga diurutkan.

2.4. Operasi Morfologi

Operasi morfologi citra merupakan suatu proses yang bertujuan untuk mengubah bentuk objek pada citra asli. Proses tersebut dapat dilakukan pada citra grayscale maupun citra biner. Jenis-jenis operasi morfologi di antaranya adalah *dilasi*, *erosi*, *closing*, dan *opening*. Seperti pada Gambar 2.6 secara berurutan, persamaan yang digunakan untuk masing-masing operasi yaitu:

$$A \oplus B \dots\dots (1)$$

$$A \ominus B \dots\dots (2)$$

$$A \bullet B = (A \oplus B) \ominus B \dots\dots (3)$$

$$A \circ B = (A \ominus B) \oplus B \dots\dots (4)$$

Gambar 2.6 Persamaan Operasi Morfologi

2.5. *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya

paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing – masing dimensi merepresentasikan fitur dari data. Algoritma KNN termasuk metode yang menggunakan algoritma *supervised*. Perbedaan antara *supervised learning* dengan *unsupervised learning* adalah pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data (Huchon et al., 1981).

Tujuan dari algoritma KNN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training *samples*. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari contoh data uji yang baru. Jarak yang digunakan adalah jarak *Euclidean Distance*.

Jarak *Euclidean* adalah jarak yang paling umum digunakan pada data numeric. Algoritma KNN merupakan algoritma yang menentukan nilai jarak pada pengujian data *testing* dengan data *training* berdasarkan nilai terkecil dari nilai ketetanggaan terdekat (Ismanto & Wardoyo, 2016).

Kedekatan didefinisikan dalam jarak matriks, seperti jarak *Euclidean*. Jarak *Euclidean* dapat dicari dengan menggunakan persamaan yang ada pada Gambar 2.7.

$$D_{xy} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$$

Gambar 2.7 Persamaan Jarak *Euclidean*

Keterangan:

D = Jarak kedekatan

x = data trainig

y = data testing

n = jumlah atribut individu antara 1 s.d. n

f = fungsi similiary atribut 1 antara kasus X dan Y

i = Atribut individu antara 1 sampai dengan n

Langkah untuk menghitung metode *K-Nearest Neighbor* antara lain:

1. Melakukan penentuan parameter K (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *euclid* (*query instance*) masing-masing objek terhadap data sampel yang diberikan menggunakan persamaan 1.
3. Mengurutkan objek tersebut ke dalam kelompok yang mempunyai jarak Euclid terkecil.
4. Mengumpulkan kategori Y (Klasifikasi Nearest Neighbor)
5. Dengan kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksi nilai *query instance* yang telah dihitung.

2.6. *Confusion Matrix*

Confusion Matrix adalah tabel yang digunakan untuk mengukur kinerja atau tingkat kebenaran dari sebuah proses klasifikasi *supervised learning* pada *machine learning*. *Confusion matrix* pada dasarnya dilakukan untuk menggambarkan lebih detail tentang jumlah data yang diklasifikasikan dengan benar maupun salah berdasarkan jumlah prediksi dari setiap data dari masing-masing kelas. *Confusion matrix* menampilkan dan membandingkan nilai aktual dan prediksi model yang digunakan untuk menghasilkan matriks evaluasi seperti *accuracy*, *precision*, *recall*, dan *f1-score* atau *f-measure*.

Tabel 2.1 *Confusion Matrix*

	Nilai Aktual		
	<i>Positive</i>	<i>Negative</i>	
Nilai Prediksi	<i>Positive</i>	TP	FP
	<i>Negative</i>	FN	TN

Seperti yang diilustrasikan pada Tabel 2.1 terdapat 4 nilai yang dihasilkan dalam tabel jika kasusnya 2 kelas, yaitu *true positive* (TP), *false positive* (FP), *false negative* (FN), dan *true negative* (TN).

Keterangan:

TP = jumlah data yang bernilai positif dan diprediksi benar positif

FP = jumlah data yang bernilai negatif tetapi diprediksi positif

FN = jumlah data yang bernilai positif tetapi diprediksi negatif

TN = jumlah data yang bernilai negatif dan diprediksi benar negatif

Akurasi untuk data 2 kelas dihitung dengan membagi jumlah data bernilai positif yang diprediksi positif dan data bernilai negatif yang diprediksi negatif dengan jumlah seluruh data dalam data set.

$$Akurasi = \frac{TP + TN}{Jumlah\ Data}$$

Data kelas yang lebih dari 3, dihitung dengan data yang bernilai positif diprediksi positif dibagi dengan jumlah data.

$$Akurasi = \frac{TP}{Jumlah\ Data}$$

2.7. Penelitian Terkait dan Matriks Penelitian

Berdasarkan penelitian yang dikembangkan oleh Yolanda Natalia dan Rudiansyah pada 2021 dengan judul “Analisis Perbandingan Bunga Meihua dan Sakura Yang Terlihat Sama Tetapi Berbeda” yang diterbitkan oleh Jurnal Bahasa, Sastra, dan Budaya, telah membahas mengenai perbedaan antara Bunga Meihua dan Sakura, dimana kedua bunga tersebut memiliki persamaan dan sulit untuk dibedakan masyarakat umum. Solusi yang diusulkan penelitian ini adalah membuat perbandingan antara bunga Meihua dan bunga Sakura dengan metode deskriptif kualitatif dengan mendeskripsikan gambaran-gambaran dari suatu fenomena yang bersifat alamiah tanpa adanya rekayasa ataupun campur tangan dari manusia. Perbandingan deskriptif dan kualitatif sudah berhasil dilakukan.

Penelitian lain juga dilakukan oleh Andi Dadang Krismawan dan Eko Hari Rachmawanto, pada tahun 2022 dengan judul “Principal Component Analysis (PCA) dan K-Nearest Neighbor (KNN) Dalam Deteksi Masker Pada Wajah” yang diterbitkan oleh ISSN. Penelitian ini dilakukan atas dasar permasalahan penggunaan masker yang diharuskan menutup hidung dan mulut untuk menghambat pertumbuhan Covid-19. Dengan adanya himbauan memakai masker dalam waktu yang lama, sistem deteksi pengenalan wajah di area umum pun terganggu dan menjadi sulit dikenali. Dengan adanya permasalahan tersebut, proses pengenalan wajah yang dilakukan sistem menjadi terhambat karena penggunaan masker yang dipakai oleh setiap orang. Berdasarkan permasalahan tersebut solusi yang ditawarkan adalah dengan membuat sistem yang dapat mengidentifikasi citra wajah seseorang meskipun menggunakan masker. Penelitian ini menggunakan Metode Principal Component Analysis (PCA) dan K-Nearest Neighbor (KNN). Data yang digunakan berupa citra wajah yang menggunakan masker dan tidak bermasker. Kemudian dilakukan tahap preprocessing dengan melakukan cropping, penghapusan latar belakang citra, segmentasi dengan menggunakan metode thresholding, ekstraksi ciri dan terakhir mengklasifikasinya menggunakan KNN. Berdasarkan hasil pengujian sistem, didapatkan akurasi sebesar 90% dengan jumlah citra sebanyak 180 citra latih dan 20 citra uji.

Tabel penelitian terkait dan matriks penelitian yang dijadikan bahan rujukan untuk melakukan penelitian ini yaitu:

Tabel 2.2 Penelitian Terkait

No	Peneliti/Tahun	Judul	Masalah Penelitian	Metode/Algoritma /Teknik/Model	State of The Art / Keterbaruan
1	Yahya, Indra Gunawan, dan Bambang Harianto (2017)	Penerapan PCA dan KNN untuk meningkatkan nilai akurasi pengenalan wajah	Proses pengenalan wajah membutuhkan algoritma dan metode yang tepat dalam meningkatkan daya jelajah dan tingkat akurasi yang akan dihasilkan	Metode/Algoritma: PCA dan K-NN	Meningkatkan nilai akurasi terhadap pengenalan wajah berdasarkan nilai jarak dengan menggunakan kombinasi algoritma K-NN dan PCA
2	Sesilia Novita R, Prihastuti Harsani, Arie Qur'ania (2018)	Penerapan KNN untuk klasifikasi anggrek berdasarkan	Banyaknya genus dan varietas tanaman anggrek yang berbeda beda diimplementasikan ke dalam	Metode/Algoritma: KNN dan <i>Euclidean</i> <i>Distance</i>	Penerapan algoritma PCA digunakan untuk mengklasifikasikan karakter morfologi daun, bentuk bunga,

		karakter morfologi daun dan bunga	sebuah aplikasi agar mudah untuk dikenali		bentuk sepal lateral, warna sepal lateral, bentuk sepal dorsal, bentuk petal, warna petal, bentuk ujung bibir, dan corak bunga
3	Husaini, Huzaeni, dan Fahmi (2018)	Algoritma PCA dalam pemrosesan sinyal electrokardiogram	PCA merupakan salah satu Teknik yang ada dalam statistic dan merupakan metode non parametrik untuk mengekstraksi informasi-informasi yang bersesuaian dari sekumpulan data yang masih diragukan	Metode/Algoritma: PCA, KNN, dan Naive Bayes	Mereduksi dimensi data-data dari sinyal EKG dengan menggunakan algoritma PCA dan akurasi klasifikasi sinyal dengan metode KNN dan Naïve Bayes

4	Andi Yulia Muniar, Pasnur, dan Kiki Ria Lestari (2020)	Penerapan algoritma KNN pada pengklasifikasian dokumen berita online	Jumlah dokumen berita online yang meningkat dengan pesat tidak dapat dikelompokkan secara efisien oleh manusia karena banyaknya data tersebut.	Metode/Algoritma: KNN	Klasifikasi dokumen berita online secara otomatis dengan menggunakan metode KNN
5	Yahya dan Winda Puspita Hidayanti (2020)	Penerapan Algoritma KNN untuk klasifikasi efektifitas penjualan vape (rokok elektrik) pada “Lombok Vape On”	“Lombok Vape On” memiliki konsumen yang cukup banyak sehingga pendapatan yang didapat tentu banyak, akan tetapi pendapatannya tidak menentu dan tidak dapat	Metode/Algoritma: KNN	Mengkasifikasikan hasil penjualan dari “Lombok Vape On” untuk mengetahui efektifitas penjualan perbulannya agar toko tersebut megetahui pasang surut pemasukan yang di dapat

			diprediksi dengan baik setiap bulannya		dengan mengimplementasikan metode KNN
6	Nofia Sari dan Resty Wulanningrum (2021)	Implementasi Algoritma KNN untuk identifikasi citra bunga anggrek	Bunga anggrek memiliki banyak jenis yang mungkin banyak masyarakat hanya mengetahui jenisnya dari warnanya saja. Banyak yang kurang mengamati tentang bunga anggrek itu sendiri	Metode/Algoritma: KNN	Implementasi algoritma KNN untuk membedakan anggrek dari kelopak bunga yang dimilikinya
7	Mainia Mayasari, Dadang Iskandar Mulyana, dan	Komparasi klasifikasi jenis tanaman rimpang menggunakan PCA,	Pemanfaatan tanaman rimpang merupakan alternatif lain sebagai obat herbal saat pandemi covid-19, tetapi masyarakat awam	Metode/Algoritma: PCA, SVM, KNN, dan <i>Decision Tree</i>	Megklasifikasikan citra tanaman rimpang dengan menggunakan metode PCA, SVM, KNN, dan <i>Decision Tree</i>

	Mesra Betty Yel (2022)	SVM, KNN dan <i>Decision Tree</i>	sering keliru saat membedakan rimpang tersebut dikarenakan memiliki kemiripan bentuk		
8	Permana Putra, Akim M H Pardede, dan Siswan Syahputra (2022)	Analisis Metode KNN dalam klasifikasi data iris bunga	Analisis metode KNN terhadap klasifikasi data iris bunga dari UCI <i>machine learning</i>	Metode/Algoritma: KNN	Mengimplementasikan metode KNN terhadap data iris bunga untuk mengetahui tingkat akurasi yang dihasilkan
9	Rini Nuraini (2022)	Klasifikasi jenis kapasitor menggunakan kombinasi	Salah satu alat elektronika yang penting adalah kapasitor. Kapasitor atau yang biasa disebut	Metode/Algoritma: KNN dan PCA	Mengembangkan sistem pengolahan citra digital untuk klasifikasi jenis transistor

		algoritma KNN dan PCA	kondensator memiliki banyak jenis. Namun, beberapa orang belum mengetahui tentang jenis-jenis kapasitor ini		dengan menerapkan metode KNN dan PCA
10	Fiqih Ismawan (2015)	Hasil Ekstraksi Algoritma <i>Principal Component Analysis</i> (PCA) untuk Pengenalan Wajah dengan Bahasa Pemrograman Java Eclipse IDE	Ekstraksi ciri menggunakan algoritma PCA untuk pengenalan wajah	Metode/Algoritma: PCA	Hasil citra wajah yang telah diekstraksi menggunakan algoritma PCA tersebut nantinya akan dibandingkan dengan citra wajah baru sebagai citra wajah yang akan dites apakah mempunyai kemiripan atau hampir mirip untuk dikenali oleh sistem

11	Yolanda Natalia dan Rudansyah (2021)	Analisis Perbandingan Bunga Meihua dan Sakura Yang Terlihat Sama Tetapi Berbeda	Perbandingan perbedaan bunga anggrek dan bunga meihua	Metode/Algoritma: Deskriptif dan Kualitatif	Membuat perbandingan dengan metode deskriptif dan kualitatif mengenai perbedaan bunga meihua dan sakura
12	Yessi Yunitasari (2020)	Klasifikasi Gambar Menggunakan Metode <i>K-Nearest Neighbor</i>	Analisis metode KNN terhadap perbedaan variasi jeruk di Indonesia	Metode/Algoritma: KNN	Mengembangkan sistem pengolahan citra digital untuk klasifikasi jenis jeruk dengan menerapkan metode KNN

13	Andi Danang Krismawan dan Eko Hari Rachmawanto (2022)	PCA dan KNN dalam deteksi masker pada wajah	Penggunaan masker mengakibatkan wajah seseorang sulit dikenali oleh sistem keamanan pada penggunaan fitur <i>face recognition</i>	Metode/Algoritma: PCA dan KNN	Mengimplementasikan metode PCA dan KNN untuk mendeteksi pengguna masker yang nantinya akan berguna bagi satgas covid-19 yang memantau masyarakat untuk selalu patuh dalam aturan penggunaan masker
14	Asahar Johar dan Delfi Yanosma (2016)	Implementasi Metode (KNN) dan (SAW) Dalam Pengambilan Keputusan Seleksi	Implementasi Metode KNN dan SAW dalam pembuatan sistem pengambilan keputusan seleksi penerimaan anggota Paskibraka	Metode/Algoritma: KNN dan SAW	Mengembangkan sistem pendukung keputusan untuk membantu pengambilan keputusan seleksi penerimaan anggota Paskibra

		Penerimaan Anggota Paskibraka			
15	Andi Danang Krismawan dan Eko Hari Rachmawanto (2022)	PCA dan KNN dalam deteksi masker pada wajah	Penggunaan masker mengakibatkan wajah seseorang sulit dikenali oleh sistem keamanan pada penggunaan fitur <i>face recognition</i>	Metode/Algoritma: PCA dan KNN	Mengimplementasikan metode PCA dan KKN untuk mendeteksi pengguna masker yang nantinya akan berguna bagi satgas covid-19 yang memantau masyarakat untuk selalu patuh dalam aturan penggunaan masker

Tabel 2.3 Matriks Penelitian

No.	Nama Peneliti	Metode						Program	Objek	Ekstraksi		
		PCA	SVM	LBP	KNN	k-Means	Lainnya			Bentuk	Warna	Tekstur
1	Yahya, Indra Gunawan, dan Bambang Harianto (2017)	√			√			-	Wajah	√		
2	Sesilia Novita R, Prihastuti				√		<i>Euclidean Distance</i>		Anggrek	√	√	

	Harsani, Arie Qur'ania (2018)											
3	Husaini, Huzaeni, dan Fahmi (2018)				√		<i>Naïve Bayes</i>	-	Sinyal	√		
4	Andi Yulia Muniar, Pasnur, dan Kiki Ria Lestari (2020)				√		<i>PHP</i>		Dokumen	√	-	-

5	Yahya dan Winda Puspita Hidayanti (2020)				√			-	Vape	-	-	-
6	Nofia Sari dan Resty Wulanningr um (2021)	√			√			Matlab	Anggrek	√	√	√
7	Mainia Mayasari, Dadang Iskandar Mulyana,	√	√		√		<i>Decision Tree</i>	-	Tanaman Rimpang	√	√	√

	dan Mesra Betty Yel (2022)											
8	Permana Putra, Akim M H Pardede, dan Siswan Syahputra (2022)				√			-	Iris Bunga	-	√	-
9	Rini Nuraini (2022)	√			√			<i>Matlab</i>	Kapasitor	√	√	-

10	Fiqih Ismawan(20 15)				√			<i>Java</i>	Wajah	√	-	-
11	Yolanda Natalia dan Rudansyah (2021)						<i>Deskriptif dan Kualitatif</i>	-	Bunga	√	-	-
12	Yessi Yunitasari (2020)				√			-	Gambar	√	√	
13	Andi Danang Krismawan dan Eko	√			√			-	Wajah	√	-	√

	Hari Rachmawan to (2022)											
14	Asahar Johar dan Delfi Yanosma (2016)				√		SAW	-	-	-	-	-
15	Andi Danang Krismawan dan Eko Hari	√			√			-	Wajah	√	-	√

	Rachmawan to (2022)											
16	Vini Rosyanti (2022)	√			√			-	Bunga	√	√	-

Berdasarkan penelitian-penelitian sebelumnya yang membahas mengenai metode *K-Nearest Neighbor* dan *Principal Component Analysis* mengimplementasikan kasus yang berbeda-beda, salah satunya adalah mendeteksi citra wajah, bunga, dokumen, transistor, dan lain sebagainya. Perbedaan penelitian ini dengan penelitian yang sebelumnya adalah akan diterapkannya klasifikasi menggunakan PCA dan KNN dalam pembuatan sistem pengolahan citra perbedaan bunga meihua dan sakura.

Hasil dari penerapan metode PCA dan KNN yang dilakukan dapat dianalisa tingkat akurasi yang dihasilkan lebih baik ataupun malah kurang dari penelitian sebelumnya.