



## Implementasi *Hybrid Intelligent Systems* Dalam Pengklasifikasian Jenis Buah *Berry*

David Agus Salim<sup>a</sup>, Agung Ramadhanu<sup>b</sup>

<sup>a,b</sup>, Magister Teknik Informatika, Universitas Putra Indonesia YPTK, Padang

### Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan jenis buah *berry* berdasarkan kualitas dan kesegarannya menggunakan metode *Principal Component Analysis* (PCA) dan *K-Nearest Neighbor* (KNN). Penelitian dilakukan dengan menggunakan 16 citra latih dan 10 citra uji, yang terdiri atas tiga jenis buah *berry*: *blackberry*, *blueberry*, dan *raspberry*. Ekstraksi fitur dilakukan menggunakan PCA untuk mengurangi dimensi data, diikuti dengan proses klasifikasi menggunakan algoritma KNN dengan jarak *Euclidean*. Hasil pengujian menunjukkan bahwa kombinasi metode PCA dan KNN mampu mencapai tingkat akurasi sebesar 100%, yang menunjukkan efektivitas model dalam mengklasifikasikan jenis buah *berry* dengan akurat. Meskipun akurasi yang tinggi telah diperoleh, diperlukan evaluasi lebih lanjut untuk memastikan kemampuan generalisasi model dan menghindari risiko *overfitting*. Penelitian ini memberikan solusi praktis untuk klasifikasi buah *berry* berdasarkan kualitas dan kesegaran dengan lebih cepat dan akurat.

**Kata Kunci:** buah *berry*, *Principal Component Analysis*, *K-Nearest Neighbor*, klasifikasi, *machine learning*.

### Abstract

*This study aims to classify berry fruit types based on quality and freshness using Principal Component Analysis (PCA) and K-Nearest Neighbor (KNN) methods. The research was conducted using 16 training images and 10 test images, consisting of three berry types: blackberry, blueberry, and raspberry. Feature extraction was performed using PCA to reduce data dimensions, followed by classification using the KNN algorithm with Euclidean distance. The testing results showed that the combination of PCA and KNN achieved an accuracy rate of 100%, indicating the model's effectiveness in accurately classifying berry fruit types. Despite the high accuracy, further evaluation is needed to ensure the model's generalization capability and to avoid overfitting. This research provides a practical solution for faster and more accurate berry fruit classification based on quality and freshness.*

**Keywords :** *berry fruit, Principal Component Analysis, K-Nearest Neighbor, classification, machine learning.*

Submitted: 03-02-2025 Approved: 23-03-2025. Published: 21-04-2025

Corresponding author's e-mail: arastyadavid@gmail.com

ISSN: Print 2722-1504 | ONLINE 2721-1002

<https://ejournal.uika-bogor.ac.id/index.php/jpg/index>

## INTRODUCTION

Buah-buahan adalah salah satu kelompok komoditas pertanian yang ada di Indonesia. Buah-buahan memiliki tingkat permintaan yang tinggi dikarenakan mengandung banyak vitamin yang baik untuk tubuh (Husna et al., 2024). Salah satu jenis buah-buahan yang cukup penting di Indonesia yaitu jenis buah berry yang juga memiliki permintaan tinggi karena kaya akan vitamin yang sangat bermanfaat bagi kesehatan tubuh (Napitu et al., 2023);(Wibi Bagas N et al., 2017). Tingginya permintaan terhadap buah-buahan *berry* di pasar domestik terlihat dari melimpahnya buah-buahan ini di berbagai pasar modern dan tradisional. Beberapa jenis *berry* yang populer di Indonesia antara lain *blackberry*, *blueberry*, dan *raspberry*. Ketiga jenis buah berry ini sangat digemari oleh masyarakat karena rasanya yang enak apalagi dibuat untuk olahan berbagai minuman. Kondisi ini mencerminkan tingginya permintaan terhadap *blackberry*, *blueberry* dan *raspberry* di kalangan masyarakat.

Seiring dengan tingginya tingkat produksi serta distribusi buah *berry*, para petani perlu memiliki kemampuan untuk mengelompokkan buah berdasarkan tingkat kesegarannya serta memperkirakan resiko pembusukan.

Pada penelitian ini akan dilakukan proses pengklasifikasian jenis buah *berry* berdasarkan kualitas dan kesegaran buah dengan menggunakan metode *Principal Component Analysis* (PCA) dan *K-Nearest Neighbor* (KNN). Manfaat dari penelitian ini yaitu dapat menghemat waktu dan memudahkan dalam pengklasifikasian sesuai jenis buah, kualitas, dan tingkat kesegaran buah.

Penelitian sebelumnya yang dilakukan (Azmi et al., 2022) membahas klasifikasi kualitas air dengan menggunakan metode *Principal Component Analysis* (PCA) untuk mengurangi dimensi fitur dan menemukan jumlah fitur optimal. Dataset yang digunakan dalam penelitian ini berasal dari situs *Kaggle*, dengan total 8.000 data yang memiliki 21 fitur pada setiap barisnya. Proses klasifikasi dimulai dengan penerapan PCA untuk mengekstraksi fitur utama, kemudian hasil ekstraksi fitur ini dijadikan input bagi dua metode klasifikasi, yaitu *K-Nearest Neighbor* (KNN) dan logistic regression.

Tujuan utama penggunaan dua metode klasifikasi berbeda adalah untuk membandingkan tingkat akurasi yang dihasilkan dari masing-masing metode. Hasil penelitian menunjukkan bahwa metode *K-Nearest Neighbor* memberikan kinerja lebih baik dibandingkan *logistic regression* dengan nilai akurasi 90,8%, presisi 90,0%, dan *recall* 91,0%. Hasil terbaik diperoleh dengan melibatkan semua fitur tanpa pengurangan signifikan dan parameter nilai  $k=9$  pada KNN. Kesimpulan dari penelitian ini menyatakan bahwa penggunaan fitur yang lengkap sangat penting dalam konteks klasifikasi kualitas air untuk mencapai hasil akurasi yang tinggi.

Penelitian terkait sebelumnya yang dilakukan (Hasym & Susilawati, 2021) membahas klasifikasi jenis ikan cupang menggunakan metode *Principal Component Analysis* (PCA) dan *K-Nearest Neighbors* (KNN). Penelitian ini bertujuan untuk membantu

pecinta ikan cupang dalam mengidentifikasi jenis ikan berdasarkan citra visual. Tiga jenis ikan cupang digunakan sebagai objek penelitian, dengan masing-masing jenis diambil sebanyak 30 citra. Total data yang digunakan terdiri dari 90 citra, yang dibagi menjadi 45 data pelatihan dan 45 data pengujian.

Hasil eksperimen menunjukkan bahwa metode KNN yang menggunakan fitur hasil reduksi dari PCA mampu menghasilkan tingkat akurasi sebesar 93,33% dalam mengenali jenis ikan cupang.

Penelitian ini menunjukkan bahwa kombinasi metode PCA dan KNN dapat memberikan hasil yang akurat dan efisien untuk klasifikasi citra ikan cupang.

Penelitian selanjutnya (Hasym & Susilawati, 2021) yang membahas klasifikasi penyakit pada daun jambu air dengan menggunakan metode *Principal Component Analysis* (PCA) dan *K-Nearest Neighbor* (KNN). Tujuan utama penelitian ini adalah mengembangkan sistem pengenalan pola yang mampu mengklasifikasikan jenis penyakit daun jambu air berdasarkan tekstur daun. Daun dipilih sebagai objek penelitian karena memiliki sifat dan ciri yang lengkap untuk membedakan jenis penyakit.

Tahapan penelitian dimulai dengan pengumpulan data citra daun jambu air yang terdiri dari 45 citra latih dan 21 citra uji. Proses ekstraksi ciri dilakukan menggunakan metode PCA untuk mengubah citra menjadi data matriks dan memperoleh nilai ciri dari setiap daun. Proses klasifikasi dilakukan menggunakan algoritma KNN dengan penghitungan jarak *Euclidean* untuk menentukan kesamaan antar data.

Hasil penelitian menunjukkan bahwa kombinasi metode PCA dan KNN mampu mengklasifikasikan penyakit daun jambu air dengan akurasi terbaik sebesar 90,48% pada nilai  $K=1$ . Sementara itu, akurasi menurun menjadi 85,71% ketika nilai  $K$  ditingkatkan menjadi 3. Penelitian ini menyimpulkan bahwa semakin besar nilai ketetanggaan  $K$  yang digunakan, semakin rendah tingkat akurasi yang diperoleh.

## **METHOD**

Pada tahap ini dilakukan dengan meninjau dan mengamati jenis buah *berry*. Data yang digunakan berupa citra jenis *berry* yaitu 16 data latih dan 10 data uji, dimana masing-masing terdiri dari 7 data latih jenis buah *blackberry*, 4 data latih *blueberry*, 5 data latih *raspberry*. Kemudian untuk data uji terdiri dari 4 buah data uji buah *blackberry*, 3 data uji *blueberry*, dan 3 data uji *raspberry*.

Berikut Gambar 1. data latih pada jenis buah *berry*:



Gambar 1. Data Latih Buah *Berry*

Adapun untuk data uji dapat dilihat pada Gambar 2. berikut:



Gambar 2. Data Uji Buah *Berry*

Selanjutnya analisis kebutuhan yang merupakan proses yang dilakukan untuk mengumpulkan informasi mengenai segala hal yang dibutuhkan dalam pengembangan sebuah sistem, sehingga sistem tersebut dapat berfungsi sesuai dengan yang diharapkan (Sriani et al., 2023).

*Principal Component Analysis (PCA)* adalah teknik yang digunakan untuk mengurangi dimensi data dengan menyederhanakan kumpulan variabel yang saling terkait, tanpa menghilangkan informasi penting yang terkandung di dalamnya (Koten, 2024); (Suhery & Ruslianto, 2017). PCA menghasilkan komponen utama yang diperoleh dari nilai eigen, yaitu *eigenvalue* dan *eigenvector*, yang dihitung berdasarkan data yang telah distandarisasi. Dengan demikian, PCA efektif dalam menangani kumpulan data yang kompleks dan mempermudah analisis data dengan tetap mempertahankan karakteristik utama dari data tersebut (Nuraini et al., 2023) ; (Aulady et al., 2023). Dengan mengurangi dimensi data, data akan lebih mudah dipahami, divisualisasi, dianalisis, serta dapat menghilangkan variabel yang tidak penting.

Langkah-langkah untuk mengurangi dimensi menggunakan algoritma PCA adalah sebagai berikut (Aulady et al., 2023); (Yusni & Zaini, 2020):

1. Pertama biarkan  $X$  menjadi matrix input untuk PCA. Dimana  $X$  merupakan data latih yang terdiri dari  $n$ -vektor dengan dimensi data  $m$ .
2. Hitung nilai mean atau rata-rata data dari setiap dimensi ( $X$ ) menggunakan persamaan (1) berikut:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

Dimana,

$n$  = Jumlah sampel data

$X_i$  = Data sampel

3. Setelah itu hitung matrix kovarians untuk mengidentifikasi korelasi menggunakan persamaan (2) berikut:

$$Cx = \frac{1}{n-1} \sum_{i=1}^n (X - \bar{X})(X_i - \bar{X})^T \quad (2)$$

Dimana,

$n$  = Jumlah sampel data

$X_i$  = Data sampel

$\bar{X}$  = Mean data

$Cx$  = Matriks kovarians

4. Hitung nilai *eigen* (*eigen value*) dan eigen vektor dari matriks kovarians yang sudah didapat menggunakan persamaan berikut:

$$CXvm = \lambda mvm \quad (3)$$

Dimana,

$CX$  = Matriks kovarians

$vm$  = *Eigen Vektor*

$\lambda m$  = *Eigen Value*

5. Langkah selanjutnya adalah mengurutkan *eigen value* dari yang terbesar ke terkecil.
6. Komponen utama (PC) adalah kumpulan eigen vektor sesuai dengan eigen value yang sudah diurutkan pada langkah 5.
7. Dimensi PC akan dikurangi berdasarkan jumlah  $K$  *eigen value* yang dipilih.
8. Data yang sampel atau uji dapat dikonversi dengan menggunakan persamaan (3) dibawah ini.

$$d = \left( (Vm)^T \left( \sum_{i=1}^n (X - \bar{X})(X_i - \bar{X})^T \right) \right) \quad (4)$$

Dimana,

$D$  = Data yang ingin dikonversi

$N$  = Jumlah sampel data

$Vm$  = *Eigen vector*

$\bar{X}$  = Mean data

$X_i$  = Data sampel

*K-Nearest Neighbor* (KNN) adalah algoritma pembelajaran terawasi yang digunakan untuk tugas klasifikasi maupun regresi. KNN bekerja dengan menentukan kelas atau label suatu data baru berdasarkan mayoritas label dari data tetangga terdekatnya, yang dihitung menggunakan jarak metrik seperti jarak *Euclidean* (Krismawan & Rachmawanto, 2022) ; (Dewi Nasien et al., 2024). Algoritma ini memiliki keunggulan berupa kemudahan interpretasi dan waktu komputasi yang relatif cepat, sehingga sering digunakan dalam berbagai kasus klasifikasi. Hasil klasifikasi ditentukan berdasarkan kelas yang paling banyak muncul di antara tetangga terdekat (Putra et al., 2023).

Dalam algoritma KNN, perhitungan jarak umumnya menggunakan *euclidean distance* sehingga untuk rumus dari KNN tersebut menggunakan rumus *euclidean distance*. Rumus algoritma KNN dapat dilihat pada persamaan (4) berikut (Aulady et al., 2023);(Sulistiani et al., 2020):

$$d(X_i, X_j) = \sqrt{\sum_{r=1}^n (ar(X_i) - ar(X_j))^2}$$

(5)

Dimana,

$d(x_i, x_j)$  = Jarak *Euclidean*

$(x_i)$  = Data peroleh ke -i

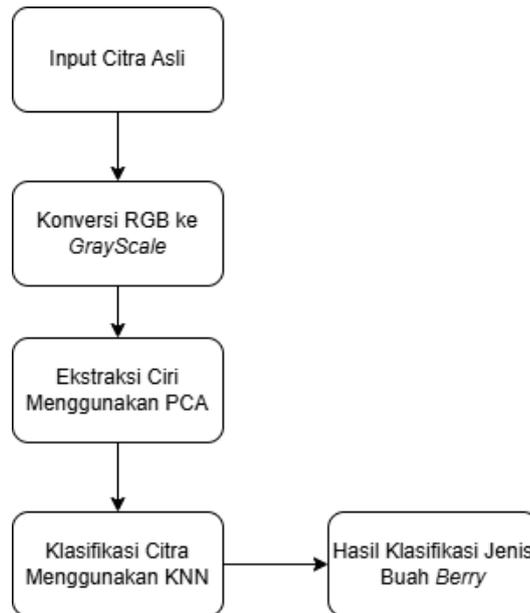
$(x_j)$  = Data peroleh ke-j

$(ar)$  = Data ke-r di dataset

I,j = 1,2,3...n

N = Jumlah data latihan

Selanjutnya dalam tahapan penelitian, dibutuhkan beberapa langkah-langkah penelitian yang dapat dilakukan agar penelitian berjalan dengan baik. Adapun langkah-langkah penelitian dapat dilihat pada Gambar 3. berikut ini.



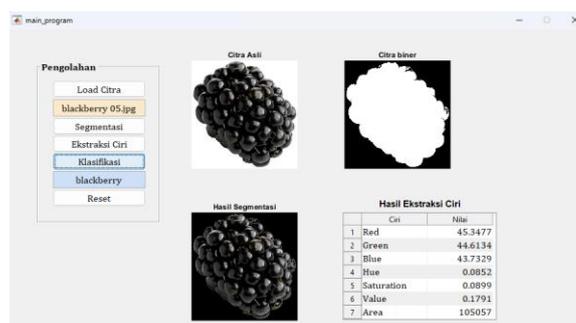
Gambar 3. Tahapan Penelitian

Tahapan dimulai dengan input citra asli yang akan mengonversi RGB ke *GrayScale*. Selanjutnya, mengekstraksi ciri menggunakan PCA. Tahap terakhir yaitu mengklasifikasikan citra menggunakan KNN dan hasil klasifikasi jenis buah *berry* ditampilkan.

## RESULTS AND DISCUSSION

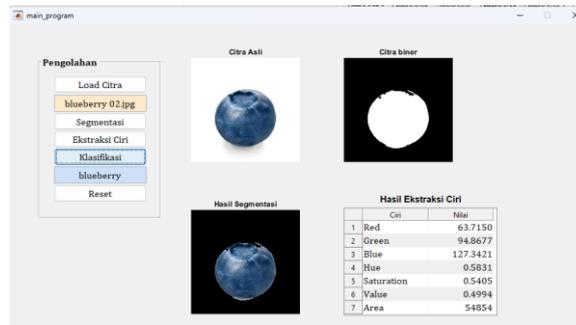
### Hasil Penelitian

Adapun proses implementasi metode PCA dan KNN dalam klasifikasi jenis buah berry dengan menggunakan sebanyak 16 citra latih dan 10 citra uji. Pengujian dilakukan dengan aplikasi GUI Matlab. Sistem dirancang terdiri dari beberapa tahap, yaitu input citra, proses segmentasi atau konversi citra *grayscale*, ekstraksi ciri, dan proses menentukan hasil klasifikasi. Proses diawali dengan menginput citra buah *berry* (*blackberry*, *blueberry*, *raspberry*) untuk dilakukan pengujian. Selanjutnya dilakukan segmentasi dan menampilkan hasil segmentasi. Dari hasil citra biner dilakukan ekstraksi ciri untuk menampilkan nilai ekstraksi ciri. Kemudian proses klasifikasi untuk menghasilkan klasifikasi jenis buah yang terlihat seperti Gambar 4. berikut.



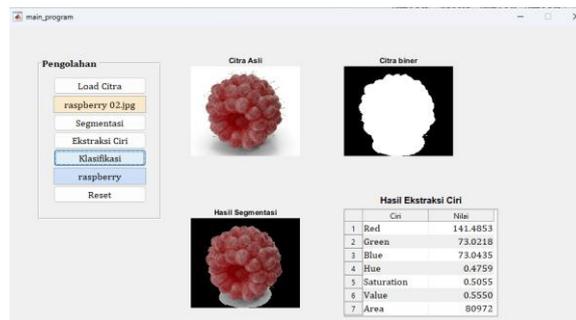
Gambar 4. klasifikasi jenis buah *blackberry*

Kemudian untuk jenis buah *blueberry* dapat dilihat pada Gambar 5. berikut.



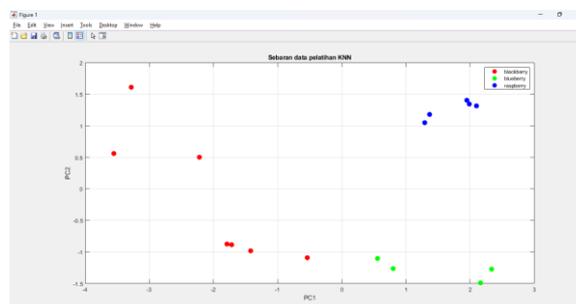
Gambar 5. klasifikasi jenis buah *blueberry*

Selanjutnya pada jenis buah *raspberry* dapat dilihat pada Gambar 6. berikut.



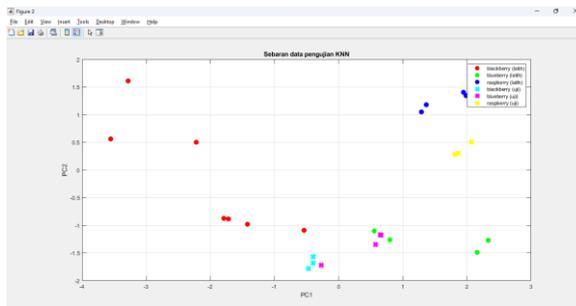
Gambar 6. klasifikasi jenis buah *raspberry*

Sesuai dengan gambar 4, 5, dan 6 diatas, proses klasifikasi berhasil diterapkan pada sistem aplikasi, dimana jenis buah *berry* berhasil diklasifikasikan antara buah *blackberry*, *blueberry*, dan *raspberry*. Hal ini menandakan bahwa klasifikasi sistem sudah benar. Adapun grafik sebaran data pelatihan KNN dapat dilihat pada Gambar 7. berikut.



Gambar 7. Grafik sebaran data pelatihan KNN

Selanjutnya untuk sebaran data pengujian KNN dapat dilihat pada Gambar 8. berikut.



Gambar 8. Grafik sebaran data pelatihan KNN

Berikut tabel hasil klasifikasi jenis buah *berry* pada program matlab.

Tabel 1. Hasil Keseluruhan Klasifikasi Data Uji

No	Nama Citra	Segmentasi	klasifikasi	Keterangan
1	blackberry 08.jpg		Blackberry	Akurat
2	blackberry 09.jpg		Blackberry	Akurat
3	blackberry 10.jpg		Blackberry	Akurat
4	blackberry 11.jpg		Blackberry	Akurat
5	blueberry 05.jpg		Blueberry	Akurat
6	blueberry 06.jpg		Blueberry	Akurat
7	blueberry 07.jpg		Blueberry	Akurat
8	raspberry 06.jpg		Raspberry	Akurat
9	raspberry 07.jpg		Raspberry	Akurat
10	raspberry 08.jpg		Raspberry	Akurat

Sesuai dengan tabel hasil pengujian klasifikasi keseluruhan citra uji diatas, langkah selanjutnya menghitung tingkat akurasi menggunakan citra uji. Berikut ini persamaannya :

$$\text{Akurasi} = \frac{\text{jumlah data uji benar}}{\text{jumlah data uji keseluruhan}} \times 100\%$$

$$\text{Akurasi} = \frac{10}{10} \times 100\% = 100\%$$

Berdasarkan perhitungan akurasi di atas, didapatkan nilai akurasi sebesar 100%, yang menunjukkan bahwa seluruh data uji berhasil diklasifikasikan dengan benar oleh model. Hal ini berarti model mampu mengenali dan mengklasifikasikan setiap citra uji tanpa kesalahan.

Namun, meskipun akurasi 100% terlihat sangat baik, perlu dilakukan analisis lebih lanjut untuk memastikan bahwa hasil ini tidak disebabkan oleh overfitting, di mana model bekerja sangat baik pada data uji tertentu tetapi tidak memiliki kemampuan generalisasi yang baik terhadap data baru.

## CONCLUSION

Berdasarkan hasil penelitian, sistem klasifikasi jenis buah *berry* menggunakan metode *Principal Component Analysis* (PCA) dan *K-Nearest Neighbor* (KNN) berhasil diimplementasikan dengan tingkat akurasi mencapai 100%.

Hasil ini menunjukkan bahwa metode PCA efektif dalam mengurangi dimensi data tanpa menghilangkan informasi penting, sementara metode KNN mampu mengklasifikasikan data.

## BIBLIOGRAPHY

- Aulady, F., Syauqy, D., & Putri, R. R. M. (2023). Sistem Klasifikasi Kualitas Air dalam Akuakultur Budidaya Ikan Lele dengan Algoritma PCA dan KNN. *Jurnal Pengembangan Teknologi ...*, 7(7), 3395–3404.  
<https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12780>
- Azmi, B. N., Hermawan, A., & Avianto, D. (2022). Analisis Pengaruh PCA Pada Klasifikasi Kualitas Air Menggunakan Algoritma K-Nearest Neighbor dan Logistic Regression. *Jurnal Sistem Dan Teknologi Informasi*, 7(2), 94–103.  
<http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/view/8190%0A>  
<http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/download/8190/4143>
- Dewi Nasien, Darwin, R., Cia, A., Leo Winata, A., Go, J., M.C, R., Charles Wijaya, R., & Charles Lo, K. (2024). Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes. *JEKIN - Jurnal Teknik Informatika*, 4(1), 10–17.  
<https://doi.org/10.58794/jekin.v4i1.640>
- Hasym, I. E., & Susilawati, I. (2021). Konvergensi Teknologi dan Sistem Informasi Klasifikasi Jenis Ikan Cupang Menggunakan Algoritma Principal Component Analysis (PCA) Dan K-Nearest Neighbors (KNN). *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 168–179.
- Husna, M., Faza, S., Lukcyhasnita, A., Yusnida, Y., & Medan, P. N. (2024). *Aplikasi diagnosa*

- penyakit tanaman buah menggunakan metode rule based reasoning. 4307(4), 1777–1781.
- Koten, I. A. C. (2024). Penerapan Metode Principal Component Analysis (Pca) Mengidentifikasi Tekstur Buah Naga. *Bina Informatika Dan Komputer (BINER)*, 2(1).  
<https://jurnal.umk.ac.id/index.php/biner/article/view/12340%0A>  
<https://jurnal.umk.ac.id/index.php/biner/article/viewFile/12340/4668>
- Krismawan, A. D., & Rachmawanto, E. H. (2022). Principal Component Analysis (PCA) dan K-Nearest Neighbor (KNN) dalam Deteksi Masker pada Wajah. *Prosiding Sains Nasional Dan Teknologi*, 12(1), 382.  
<https://doi.org/10.36499/psnst.v12i1.7066>
- Napitu, S., Paramita Panjaitan, R., Nulhakim, P. A., & Khalik Lubis, M. (2023). Klasifikasi Buah Jeruk Segar dan Busuk Berdasarkan RGB dan HSV Menggunakan Metode KNN. *Jurnal SAINTEKOM*, 13(2), 214–221.  
<https://doi.org/10.33020/saintekom.v13i2.420>
- Nuraini, R., Wibowo, A., Warsito, B., Syafei, W. A., & Jaya, I. (2023). Combination of K-NN and PCA Algorithms on Image Classification of Fish Species. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 7(5), 1026–1032.  
<https://doi.org/10.29207/resti.v7i5.5178>
- Putra, F. A.-I. A., Sulaksono, A. G., Utomo, L. T., & Khamdani, A. R. (2023). Klasifikasi Buah dan Sayur Menggunakan Fitur Ekstraksi HOG dan Metode KNN. *Jurnal Informatika Polinema*, 10(1), 45–52.  
<https://doi.org/10.33795/jip.v10i1.1433>
- Sriani, Supriyandi, Furqan, M., & Fadilla Rischa, W. (2023). Pengenalan Pola Penyakit Daun Jambu Air Menggunakan Metode PCA dan KNN. *Jurnal Jaringan Sistem Informasi Robotik (JSR)*, 7(2), 158–163.  
<http://ojsamik.amikmitragama.ac.id>
- Suhery, C., & Ruslianto, I. (2017). Identifikasi Wajah Manusia untuk Sistem Monitoring Kehadiran Perkuliahan menggunakan Ekstraksi Fitur Principal Component Analysis (PCA). *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 3(1), 9.  
<https://doi.org/10.26418/jp.v3i1.19792>
- Sulistiani, H., Darwanto, I., & Ahmad, I. (2020). Penerapan Metode Case Based Reasoning dan K-Nearest Neighbor untuk Diagnosa Penyakit dan Hama pada Tanaman Karet. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 6(1), 23.  
<https://doi.org/10.26418/jp.v6i1.37256>
- Wibi Bagas N, H., Mailoa, E., & Purnomo, H. D. (2017). Deteksi Buah untuk Klasifikasi Berdasarkan Jenis dengan Algoritma CNN Berbasis YOLOv3. *Masa Berlaku Mulai*, 1(3), 476–481.
- Yusni, R. mulyadi, & Zaini. (2020). Identifikasi Pengenalan Wajah Perokok Menggunakan Metode Principal Component Analysis. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 892–898.  
<https://doi.org/10.29207/resti.v4i5.2272>