

# ***K-Means Clustering Dalam Proses Pengenalan Daun Untuk Mengidentifikasi Tanaman Herbal***

## ***K-means Clustering In Leaf Recognition Process To Identify Herbal Plants***

1<sup>st</sup> M. Ghalib Riswandi  
*Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia*  
ghalibriswandi@telkomunive  
rsity.ac.id

2<sup>nd</sup> Anggunmeka Luhur  
Prasasti  
*Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia*  
anggunmeka@telkomuniver  
sity.ac.id

3<sup>rd</sup> Marisa W. Paryasto  
*Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia*  
marisaparyasto@telkomunive  
rsity.ac.id

**Abstrak**—Tanaman herbal adalah tanaman yang biasanya dapat dimanfaatkan pada bagian daun. Mengenai tanaman herbal dimana warna dan bentuk dari suatu daun tanaman herbal yang tergolong mirip maka akan menjadi sulit untuk membedakan jenis dan nama daun serta khasiat dari daun tanaman herbal tersebut. Image processing menjadi salah satu pilihan untuk mempermudah manusia untuk membedakan nama, jenis dan manfaat daun tanaman herbal tersebut. Pada tugas akhir ini digunakan sebanyak 5 jenis daun tanaman herbal dan digunakan GLCM sebagai ekstraksi fitur serta K-Means Clustering sebagai klasifikasi daun.

**Kata kunci** — *Image processing, Gray Level Co-occurrence Matriks, K-Means Clustering.*

**Abstract**—Herbal plants are plants that can usually be used on the leaves. Regarding herbal plants where the color and shape of a leaf plant of herbal plants that are classified as similar, it will be difficult to distinguish the type and name of the leaf and the efficacy of the leaves of the herbal plant. Image processing is an option to make it easier for humans to distinguish the names, types and benefits of the leaves of these herbal plants. In this final project, 5 types of leaves of herbal plants were used and GLCM as feature extraction and K-Means Clustering as leaf classification.

**Keywords**— *Image processing, Gray Level Co-occurrence Matriks, K-Means Clustering.*

### I. PENDAHULUAN

Banyak teknologi yang diciptakan dan digunakan oleh manusia untuk mempermudah pekerjaannya di dalam segala bidang seperti contohnya pembuatan alat yang menggunakan teknologi atau metode *Image Processing* untuk menentukan dan mengklasifikasikan berbagai benda [1]. Berkaitan dengan hal tersebut maka dapat dibuat suatu alat yang berhubungan dengan *Image Processing* tentang Menentukan dan Mengklasifikasikan jenis tanaman herbal tertentu berdasarkan jenis daun. Pembuatan alat ini didasari oleh masalah yang akan muncul pada saat ini mengenai klasifikasi daun pada tanaman herbal dimana warna dan ukuran serta bentuk yang terlihat mirip satu dengan yang lainnya, Pemanfaatan sederhana dari *Image Processing* memudahkan bagi pengguna yang akan memanfaatkan alat ini sebagai indikator pemisah antara jenis daun pada tanaman herbal berdasarkan jenisnya. ekstraksi fitur pada objek yang akan diklasifikasi memiliki fungsi untuk meningkatkan akurasi dan mengurangi kesalahan [2].

### II. KAJIAN PUSTAKA

#### A. Tanaman Herbal

Kebanyakan tanaman herbal yang biasanya diambil manfaat atau khasiatnya adalah berdasarkan daun atau akar, namun tanaman herbal yang berada di Indonesia

kebanyakan dimanfaatkan berdasarkan daun. Daun tanaman herbal berjumlah 5 jenis yang terdiri dari, daun mengkudu (*Morinda citrifolia* L), daun sirih (*Piper betle* L), daun teh (*Camellia Sinensis*), daun betadine (*Jatropha multifida* In) dan daun kersen (*Muntingin calabura* L). Masing masing dari daun tersebut difoto dan dijadikan dataset sebanyak 5 jenis daun tanaman herbal, sedangkan setiap jenisnya terdiri dari 50 foto bahkan lebih, pembuatan dataset secara

manual dikarenakan sulitnya mencari dataset daun tanaman herbal yang tidak tersedia di banyak tempat, sedangkan di Negara Indonesia sendiri sangat banyak jenis tanaman herbal berdasarkan daunnya. Pengenalan daun dapat juga dilihat berdasarkan dari pola suatu tulang daun yang dapat dilakukan segmentasi dengan Curvature points Methode, metode ini cukup baik dalam segmentasi untuk pelacakan lokasi tulang daun [3].



GAMBAR 1  
DAUN TANAMAN HERBAL.

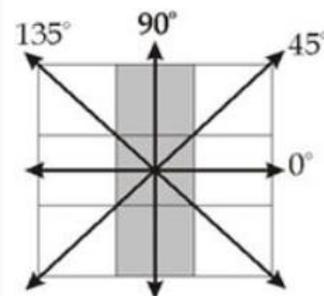
#### B. Image Processing

*Image Processing* adalah suatu pemrosesan sinyal dengan gambar sebagai input serta gambar lain yang berupa output menggunakan teknik tertentu. *Image Processing* yang baik adalah *image* yang memiliki kualitas gambar yang tinggi serta sesuai dengan gambar aslinya. *Image processing* menggunakan komputer adalah dengan tujuan mengolah suatu gambar menjadi gambar yang memiliki kualitas baik dari sebelumnya atau pengelompokan gambar berdasarkan jenisnya [4]. Pada penelitian ini, masih terdapat beberapa kekurangan karena teknik pengambilan gambar yang belum optimal. Hal ini dapat disebabkan oleh pengambilan gambar yang buruk, jarak pemotretan yang tidak memadai, dan kondisi pencahayaan yang minim. Salah satu solusi yang dapat dilakukan untuk mengatasi kekurangan pada citra yang diambil adalah dengan menggunakan metode *image enhancement* [5]. Penggunaan utama *Image Enhancement* di setiap bidang adalah untuk mendapatkan informasi yang jelas. Karena informasi ini merupakan sesuatu yang sangat penting dalam kehidupan sehari-hari. Hal ini juga untuk meminimalisir kesalahpahaman terhadap informasi yang kita terima [6].

#### C. Gray Level Co-occurrence Matrix

GLCM *Gray Level Coocurence Matrix* atau GLCM adalah Teknik suatu analisis tekstur pada citra atau gambar. GLCM merepresentasikan hubungan antara 2 citra

yang bersebelahan (*neighboring pixels*) yang mempunyai intensitas warna keabuan (*grayscale intensity*) sudut dan jarak. GLCM memiliki jumlah sudut sebanyak 8 buah diantaranya adalah sudut  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $180^\circ$ ,  $225^\circ$ ,  $270^\circ$ , atau  $315^\circ$ . Pada penelitian kali ini, penulis hanya menggunakan sudut ideal yang dimiliki GLCM yaitu sudut  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ .



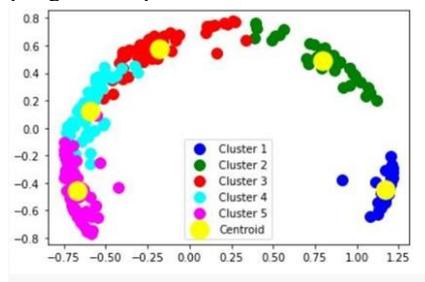
GAMBAR 2  
GRAY LEVEL CO-OCCURENCE MATRIX.

#### D. K-Means Clustering

*K-Means Clustering* atau algoritma *K-Means* adalah algoritma pengelompokan yang digunakan untuk mensegmentasi atau mempartisi sekumpulan objek data yang akan dijadikan subset dan dapat disebut cluster. Data-data atau objek yang ada pada suatu cluster mempunyai karakteristik yang mirip antara objek satu dengan objek lainnya. Segmentasi suatu objek tidak dilakukan secara manual, tetapi dengan mengimplementasikan algoritma *K-Means*

Clustering untuk mendapatkan kelompok objek atau data yang diinginkan. Algoritma clustering ini banyak diaplikasikan dalam ilmu kecerdasan buatan, pengenalan pola

gambar dan lain lain yang berhubungan dengan banyak grup, salah satunya adalah pengelompokan [7].



GAMBAR 3  
K-MEANS CLUSTERING

Rumus dalam *K-Means Clustering* untuk mencari pusat suatu cluster dapat dituliskan sebagai berikut:

$$d = |x - y| = \sqrt{\sum_{i=1}^h (x_i - y_i)^2} \tag{2.1}$$

**Keterangan:**

X = Pusat *Cluster*

Y = Data Centroid

Rumus dalam *K-Means Clustering* untuk mencari centroid:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} X_q \tag{2.2}$$

**Keterangan:**

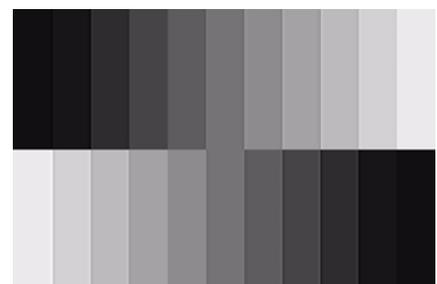
$\mu_k$  = Centroid point of the *K-Cluster*

$N_k$  = jumlah data didalam *K-Cluster*

$X_q$  = q-data didalam *K-Cluster*

E. Citra Grayscale

Citra atau gambar yang dapat ditunjukkan oleh citra *grayscale* terdiri dari warna abu-abu, sedangkan diantaranya adalah warna hitam sebagai bagian yang terlemah dan pada sisi lainnya adalah warna putih sebagai intensitas yang kuat. Citra *Grayscale* memiliki intensitas tingkat keabuan yang disebut (*gray level*). Untuk rentang yang biasa digunakan direpresentasikan: 0.0 (hitam) hingga 1.0 (putih) sedangkan *Grayscale* adadiantara warna hitam dan putih.



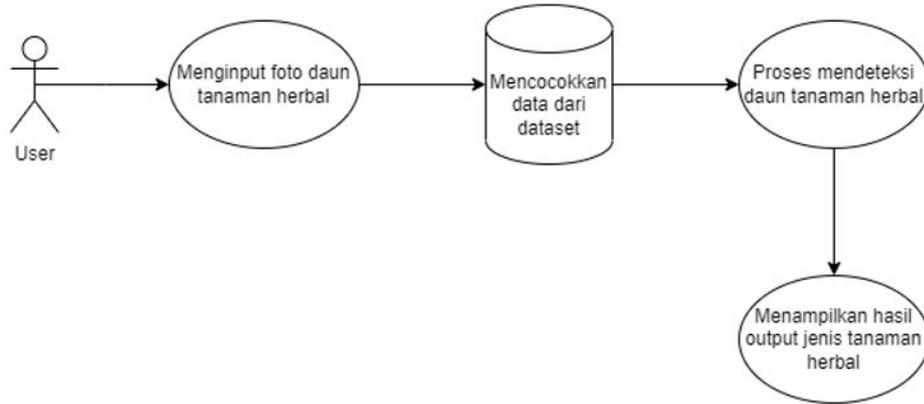
GAMBAR 4  
CITRA GRAYSCALE

III. METODE

A. Desain Sistem

Gambar diatas merupakan gambaran umum sistem yang dibuat berdasarkan garis besar sistemnya. Dengan memasukkan foto daun tanaman herbal yang akan diidentifikasi

menggunakan *machine learning* supaya dapat dideteksi daun tanaman herbal berdasarkan jenisnya, nama tanaman herbal, nama ilmiah daun tanaman herbal serta manfaatnya maka perlu mencocokkan data dengan dataset yang telah disediakan.

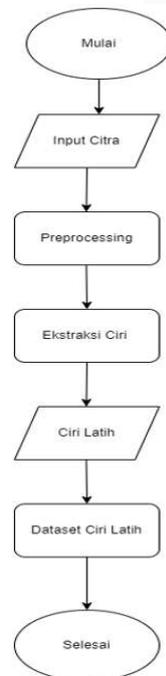


GAMBAR 5  
DESAIN SISTEM.

B. Diagram Alir Sistem Training

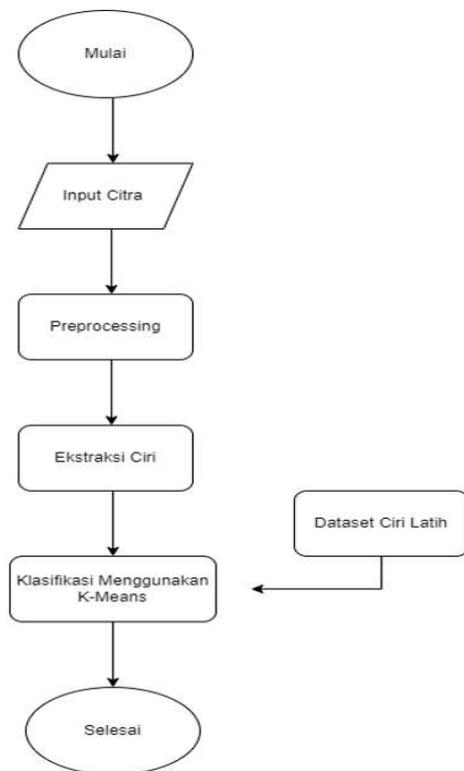
Pada gambar diatas merupakan proses *training* data dan proses mendeteksi daun tanaman herbal. Proses ini dilakukan dengan menginput citra yang diambil melalui kamera selanjutnya akan dilakukan proses

*preprocessing* data dimana citra akan dilakukan perubahan warna citra dari citra RGB ke *grayscale*.



GAMBAR 6  
DIAGRAM ALIR SISTEM TRAINING.

C. Diagram Alir Sistem Testing



GAMBAR 7  
DIAGRAM ALIR SISTEM TESTING.

Diagram diatas proses ini dilakukan dengan menginput citra yang diambil melalui camera atau menginput gambar yang sudah ada selanjutnya akan dilakukan *processing* dimana citra akan dilakukan perubahan warna citra dari citra RGB ke *grayscale*, pada *OpenCV* untuk mengubah RGB ke *grayscale* menggunakan perintah `COLOR_BGR2GRAY` data dan akan dilakukan ekstraksi fitur menggunakan Gray Level Co-Occurrence Matriks. menggunakan algoritma *K-means* dimana *K-means* merupakan salah satu algoritma yang bersifat unsupervised learning. *K-Means* memiliki fungsi untuk mengelompokkan data kedalam data cluster.

IV. HASIL DAN PEMBAHASAN

A. Pengujian Fitur Ekstraksi Ciri

Pada pengujian ini dilakukan analisis pengujian GLCM yang bertujuan untuk meningkatkan performansi sistem untuk presentase peningkatkan akurasi klasifikasi. Pengujian ini menggunakan dataset sebanyak 51 buah dengan jenis daun yang otomatis dipilih secara acak. Berikut hasil pengujian menggunakan 4 fitur yaitu *Dissimilarity*, *Correlation*, *Homogeneity* dan *Contrast* pada GLCM.

TABEL 1  
HASIL PENGUJIAN 4 PARAMETER.

Kombinasi 4 parameter	Akurasi
Dissimilarity, Correlation, Homogeneity, Contrast	90%

Dengan menggunakan 4 fitur pada GLCM ini, dapat diperoleh akurasi paling besar yaitu sebesar 90% dengan menggunakan algoritma *K-Means Clustering*.

B. Pengujian Jarak Terhadap Sudut

Pada pengujian ini dilakukan perubahan parameter distance pada Gray Level Co-Occurrence (GLCM) terhadap sudut yang memberikan nilai akurasi paling tinggi. Berikut adalah hasil pengujian dengan menggunakan empat fitur contrast, correlation, homogeneity dan dissimilarity pada GLCM.

TABEL 2  
PENGUJIAN PARAMETER JARAK TERHADAP SUDUT.

Sudut Orientasi	Pengujian Parameter Jarak Terhadap Sudut				
	1	2	3	4	5
0°	88%	90%	86%	90%	90%
45°	88%	90%	86%	90%	90%
90°	88%	90%	86%	90%	90%

135°	88%	90%	86%	90%	90%
------	-----	-----	-----	-----	-----

Pengujian parameter jarak terhadap sudut GLCM yang terdapat empat jenis sudut atau arah ini bertujuan untuk mencari nilai akurasi yang terbaik dari sudut **0°, 45°, 90°, 135°** Parameter yang diubah adalah jarak 1, 2, 3, 4 dan 5. Dari hasil pada table diatas, dapat disimpulkan bahwa sudut 1, 4 dan 5 adalah parameter paling bagus hasilnya dengan nilai 90% jika dibandingkan dengan sudut 1 dan 3 yang hasilnya adalah 88% dan 86%.

Pada Pengujian masing-masing jenis daun dapat dibuktikan bahwa menggunakan algoritma K-Means Clustering memiliki tingkat identifikasi sangat cepat.

Dibuktikan dengan rata-rata pengujian gambar hanya 0.34 detik waktu komputasi. Identifikasi gambar pada bagian akurasi tingkat kecocokan data, K-Means Clustering memiliki nilai rata-rata presentase sebesar 73.72%.

### C. Pengujian Parameter SVM Terhadap *K-Means Clustering*

Dilakukan pengujian dua parameter kernel pada Support Vector Machine (SVM) terhadap *K-Means Clustering*, pada parameter SVM yang akan diuji yaitu kernel rbf dan polynomial (poly). Berikut merupakan hasil dari pengujian yang dilakukan:

TABEL 3  
PENGUJIAN KERNEL RBF TERHADAP K-MEANS CLUSTERING

Parameter SVM		Kernel	Pengujian Kernel RBF Terhadap <i>K-Means</i>				Akurasi
C	Gamma		n_cluster	init	n-init	max_iter	
0.1	1	Rbf	5	random	10	300	20%
1	1	Rbf	5	random	10	300	20%
10	1	Rbf	5	random	10	300	20%
100	1	Rbf	5	random	10	300	20%
1000	1	Rbf	5	random	10	300	20%
10000	1	Rbf	5	random	10	300	20%
0.1	0.1	Rbf	5	random	10	300	20%
0.1	0.01	Rbf	5	random	10	300	20%
0.1	0.001	Rbf	5	random	10	300	20%
0.1	0.0001	Rbf	5	random	10	300	20%
0.1	0.00001	Rbf	5	random	10	300	20%

Pada table diatas dapat dilihat bahwa pengujian pada parameter SVM yang diubah nilainya adalah nilai C dan Gamma, pada

nilai C dan Gamma diperoleh nilai yang sama yaitu hanya 20% pada kernel RBF.

TABEL 4  
PENGUJIAN KERNEL POLY TERHADAP K-MEANS CLUSTERING

Parameter SVM		Kernel	Pengujian Kernel POLY Terhadap <i>K-Means</i>				Akurasi
C	Gamma		n_cluster	init	n-init	max_iter	
0.1	1	Poly	5	random	10	300	86%
1	1	Poly	5	random	10	300	86%
10	1	Poly	5	random	10	300	86%
100	1	Poly	5	random	10	300	86%
1000	1	Poly	5	random	10	300	86%
10000	1	Poly	5	random	10	300	86%
0.1	0.1	Poly	5	random	10	300	86%
0.1	0.01	Poly	5	random	10	300	88%
0.1	0.001	Poly	5	random	10	300	88%

0.1	0.0001	Poly	5	random	10	300	86%
0.1	0.00001	Poly	5	random	10	300	75%

Sedangkan pengujian dengan parameter SVM yang sama nilainya, Kernel "POLY" memberikan hasil yang cukup bagus dengan

angka terbesar yaitu 88% dengan parameter C=0.1, parameter Gamma=0.01 dan 0.001.

#### D. Pengujian Hasil Komputasi

TABEL 5  
WAKTU KOMPUTASI SISTEM

Jenis Daun	Tingkat Kecocokan (%)	Waktu Pemrosesan (detik)
Betadine	84.58%	0.295
Kersen	70.26%	0.284
Mengkudu	78.77%	0.310
Sirih	76.57%	0.404
Teh	58.40%	0.409
Rata-rata	73.71%	0.340

Pada pengujian yang dilakukan untuk mendapatkan hasil pemrosesan data melalui website dengan melakukan percobaan sepuluh gambar daun per jenisnya, mendapatkan hasil rata rata tingkat kecocokan sebesar 73.71% dengan rata rata waktu pemrosesan sebesar 0.340 detik.

#### V. KESIMPULAN

*K-Means Clustering* sangat baik pada bagian kecepatan waktu pengujian, hal ini disebabkan karena *K-Means Clustering* memiliki fungsi yang kompleks dan sederhana, namun pada bagian akurasi tingkat kecocokan data, *K-Means Clustering* memperoleh nilai presentase sebesar 73.72 % dan data rata-rata pengujian gambar hanya dengan 0.34 detik waktu pemrosesan pada *K-Means Clustering*.

#### REFERENSI

- [1] M. F. Naufal, "Analisis Perbandingan Algoritma Svm, Knn, Dan Cnn," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 8, p. 311, 2019.
- [2] PRASASTI, A. L., IRAWAN, B., FAJRI, S. E., RENDIKA, A., & HADIYOSO, S. (2020). Perbandingan Ekstraksi Fitur dan Proses Matching pada Autentikasi Sidik Jari Manusia. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 8(1), 95.
- [3] Prasasti, A. L., Mengko, R. K. W., & Adiprawita, W. (2015). Vein tracking using 880nm near infrared and CMOS sensor with maximum curvature points segmentation. In 7th WACBE World Congress on Bioengineering 2015 (pp. 206-209). Springer, Cham.
- [4] Putra, R. D., Purboyo, T. W., & Prasasti, L. A. (2017). A Review of Image Enhancement Methods. *International Journal of Applied Engineering Research*, 12(23), 13596-13603.
- [5] Aziz, M. N., Purboyo, T. W., & Prasasti, A. L. (2017). A survey on the implementation of image enhancement. *Int. J. Appl. Eng. Res*, 12(21), 11451-11459.
- [6] Z. Wang Y. Zhou, "Anomaly Detection by Using Streaming K-Means and Batch K-Means," *IEEE International Conference on Big Data Analytics*, p. 11, 2020.
- [7] N. Noriko, "Potensi Daun Teh (*Camellia sinensis*) dan Daun Anting," *Jurnal Al-Azhar Indonesia Seri Sains Dan Teknologi*, 2, P. 104, 2013.
- [8] Ilkafah, "Daun Kersen (*Muntingia calabura L.*) Sebagai Alternatif Terapi," *Pharmacy Medical Journal*, 1, p. 33, 2018.
- [9] Y. Liana Y. A. Utama, "Efektifitas

- pemberian ekstrak daun betadine (*Jatropha multifida* Linn) terhadap ketebalan jaringan granulasi dan jarak tepi luka pada penyembuhan luka sayat tikus putih (*Rattus norvegicus*),” *JKK*, 5, p. 114, 2018.
- [10] R. M. Prakash, G.P.Saraswathy, G.Ramalakshmi, K.H.Mangaleswari T.Kaviya, “Detection of Leaf Diseases and Classification using Digital Image Processing, ” *International Conference on Innovations in Information, Embedded and Communication Systems (ICIECS)*, p. 978, 2017.
- [11] R. Indrawan, N. Muhammad, L. P. Anggunmeka, “Implementation of K-Means Clustering Algorithm based on Rice Productivity Level in Subdistrict Area,” *Research Journal of Applied Sciences*, 14, p. 110, 2019.
- [12] Microsoft, “ Visual Studio, ” Microsoft, 29 04 2021. Available: <https://code.visualstudio.com/docs/setup/setup-overview>. 10 12 2021].
- [13] Parapat, I. Monika, Furqon, M. Tanzil Sutrisno, “Penerapan Metode Support Vector Machine (SVM) pada klasifikasi Penyimpangan Tumbuh Kembang Anak,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2, p.3163, 2018.
- [14] I. S. G. d. S. A. P. R.A. Rizal, “Klasifikasi Wajah Menggunakan Support Vector Machine (SVM), ” *Riset dan E-Jurnal Manajemen Informatika Komputer*, 3, 2019.
- [15] T. K.Akbar Nur Syahrudin, “ Input Dan Output Pada Bahasa Pemrograman Python,,” *Jurnal dasar Pemrograman Python STMI K,p.1.*, p. 1, 2018.
- [16] S. v. d. Walt, J. L. Schonberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager E. G. a. T. Yu., “scikit-image: image processing in Python,” *PeerJ*,p.453, 2014.
- [17] M. F. Naufal, “Analisis Perbandingan Algoritma Svm, Knn, Dan Cnn ,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 8, p. 311, 2019.
- [18] Bustanussalam, D. Apriasi, E. Suhardi D. Jaenudin, “ Efektivitas Antibakteri Ekstrak Daun Sirih (*Piper betle* Linn) Terhadap *Staphylococcus aureus* ATCC 25923,” *Fitofarmaka*, 5, p. 58, 2015.
- [19] Halimah, D. M. Suci I. Wijayanti, “ Studi Potensi Penggunaan Daun Mengkudu (*Morinda citrifolia* L.) sebagai Bahan Antibakteri *Escherichia coli* dan *Salmonella typhimurium*,” *Jurnal Ilmu Pertanian Indonesia (JIPI)*, 24, p. 58, 2019.