

IMPLEMENTASI GLCM DAN ALGORITMA NAÏVE BAYES DALAM KLASIFIKASI JENIS BUNGA ANGGREK

Ivan Seth Manuel¹, Iin Ernawati²

Program Studi Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia.

ivan.seth77@gmail.com¹, iinerti@gmail.com²

Abstrak. Berdasarkan banyaknya spesies dan kemiripan yang dimiliki pada setiap jenis bunga anggrek membuat masyarakat dan para pembudi daya sulit dalam membedakan serta membutuhkan waktu yang cukup lama dalam melakukan pengklasifikasian. Penelitian ini menggunakan algoritma Naïve Bayes dalam proses klasifikasi. Data yang digunakan bersumber dari citra bunga anggrek yang diambil secara langsung menggunakan kamera. Menggunakan nilai ekstraksi ciri GLCM sebagai data latih dan parameter dalam melakukan pengklasifikasian Naïve Bayes. Berdasarkan hasil pengujian, mendapatkan nilai akurasi sebesar 61,1% hal ini dikarenakan proses kombinasi yang dilakukan menghasilkan interval jarak nilai ekstraksi ciri yang berdekatan antar kelasnya sehingga pengklasifikasian Naïve Bayes kurang berjalan dengan optimal.

Kata Kunci: Bunga Anggrek, Klasifikasi, GLCM, Naïve Bayes

1 Pendahuluan

Berdasarkan banyaknya spesies anggrek dan kemiripan yang dimiliki pada setiap jenis anggrek membuat masyarakat sulit membedakan dan perlunya mengklasifikasikan spesies anggrek satu dengan yang lain. Anggrek memiliki banyak keunikan yang menjadi daya tarik bagi para penggemar tanaman hias, bagian yang menjadi daya tarik pada Anggrek terletak pada bunga dan warna bunganya yang beragam.

Klasifikasi bunga yang digunakan saat ini berdasarkan mahkota bunga dan susunan bunga, pengklasifikasian yang dilakukan dengan cara manual dengan melihat langsung bunga yang akan diklasifikasikan. Dalam prakteknya jika dilakukan dalam jumlah besar akan memerlukan tenaga yang besar dan kurang efektif serta memerlukan waktu yang cukup lama dalam pengklasifikasian[1].

Pada penelitian yang dilakukan oleh Danar Putra Pamungkas dalam melakukan identifikasi bunga anggrek menggunakan metode GLCM sebagai ekstraksi ciri, tingkat keberhasilan yang diperoleh sebesar 80% [2]. Dengan penelitian menggunakan metode GLCM dan Naïve Bayes dalam melakukan pengelompokan buah jeruk yang dilakukan oleh Haba dan Pelangi memperoleh hasil pengujian sebesar 82% [3].

Berdasarkan penelitan – penelitian yang telah dilakukan sebelumnya dengan menggunakan metode GLCM dan Naïve Bayes dalam melakukan identifikasi dan pengelompokan. Oleh karena itu penelitian ini dilakukan untuk dapat mengklasifikasikan bunga anggrek secara komputerisasi berdasarkan ekstraksi ciri yang menjadi karakteristik pada bunga anggrek sehingga dapat membedakan jenis bunga anggrek yang satu dengan jenis

bunga anggrek yang lain secara efektif dan efisien. Klasifikasi jenis bunga anggrek yang dilakukan menggunakan metode *Gray Level Co – Occurrence Matrix* (GLCM), yang digunakan sebagai metode untuk ekstraksi fitur sebagai analisa tekstur pada citra dan metode algoritma *Naïve Bayes* sebagai pengklasifikasi citra yang menggunakan perhitungan probabilitas kemiripan kasus yang lama terhadap objek tertentu.

2 Landasan Teori

2.1 Image Processing

Pengolahan Citra (*Image Processing*) adalah sebuah sistem yang dapat memproses berupa citra dan hasilnya juga berupa citra [3]. Pengolahan citra ini dilakukan untuk memperbaiki kualitas citra dengan kecepatan proses komputer, serta memungkinkan manusia dapat mengambil informasi dari citra. Citra harus direpresentasikan secara numerik dengan nilai – nilai diskrit agar dapat di olah dengan komputer digital [4].

Metode Deteksi Tepi Canny. Deteksi Tepi Canny merupakan algoritma deteksi tepi yang banyak digunakan karena dinilai sebagai algoritma deteksi tepi yang paling optimal [5]. Berikut merupakan tahapan yang dilakukan dari deteksi tepi *canny*:

1. Menghaluskan citra dengan menggunakan *Filter Gaussian*. Penghapusan derau dengan memakai *filter Gaussian blur* yaitu operasi konvolusi yang dilakukan operasi perkalian antara matriks kernel dengan matriks gambar asli. Hasil dari penghalusan citra ini akan di gunakan untuk penghitungan potensi gradien.
2. Menghitung gradien citra
Hal – hal yang dilakukan yaitu magnitudi (*Magnitude/edge strength*) dan orientasi (*direction/orientation*). Gradien yang ada pada setiap piksel citra telah diperhalus dengan menerapkan operator *Sobel* sebagai magnitude dari gradien:

$$sx = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 2 \end{bmatrix} \quad sy = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Kemudian dilakukan penguatan tepi. Dilakukan dengan operator Gaussian. Gradient citra dihitung melalui rumus:

$$M = \sqrt{sx^2 + sy^2}$$

Kemudian melakukan penghitungan arah tepi menggunakan persamaan:

$$\theta \arctan = \left(\frac{Gy}{Gx} \right)$$

Setelah itu melakukan penghubungan arah tepi kedalam salah satu kategori dari empat arah derajat berdasarkan ketentuan.

3. *Non-Maximum Suppression*. Penghilangan non-maksimum dilakukan pencarian lokasi tepi dari citra, pada sepanjang tepi berdasarkan arah tepi dan menghilang piksel yang tidak dianggap sebagai tepi menjadi 0.
4. *Hysteresis Thresholding*
Menerapkan *threshold* untuk menentukan tiap piksel yang dikategorikan dalam tepi atau tidak. Pada tahap ini menggunakan nilai *threshold* bawah dan *threshold* atas. Jika nilai pada piksel lebih besar dari *threshold* bawah dan masih terhubung dengan *threshold* atas maka akan di tentukan sebagai tepi.

Metode Grey Level Co – Occurrence Matrix (GLCM). *Grey Level Co – Occurrence Matrix (GLCM)* merupakan teknik untuk mendapatkan statistic orde kedua dengan menghitung probabilitas hubungan kedekatan antara dua piksel dalam jarak spesifik dan orientasi derajat [6]. Sudut yang digunakan adalah 0, 45, 90, 135 dan jarak yang digunakan tergantung dengan ketetanggaan antara piksel. Metode GLCM memiliki beberapa nilai fitur yang dihasilkan yaitu:

a. *Contrast*

Merupakan salah satu variable yang terdapat dalam GLCM yang berfungsi dalam mengukur intensitas kontras keabuan pada citra.

$$Con = \sum_i \sum_j (i - j)^2 p_{(i,j)} \quad (1)$$

Keterangan:

Con : Nilai *contrast*

j : Indeks kolom pada matrik

i : Indeks baris pada matrik

$P(i, j)$: Probabilitas kolom(*i, j*)

b. *Correlation*

Merupakan variable yang digunakan untuk mengukur tingkat keterkaitan linear dari derajat dengan derajat keabuan citra

$$Cor = \sum_i \sum_j \frac{(i - \mu_i)(j - \mu_j)p_{(i,j)}}{\sigma_i \sigma_j} \quad (2)$$

Dimana:

$$\mu_i = \sum_i \sum_j i p_{(i,j)} \quad \sigma_i = \sqrt{\sum_i \sum_j (i - \mu_i)^2 p_{(i,j)}}$$

$$\mu_j = \sum_i \sum_j j p_{(i,j)} \quad \sigma_j = \sqrt{\sum_i \sum_j (j - \mu_j)^2 p_{(i,j)}}$$

Keterangan:

Cor : Nilai *correlation*

μ_i : Jumlah nilai baris matrik GLCM normalisasi

μ_j : Jumlah nilai kolom matrik GLCM normalisasi

σ_i : Nilai baris dari total μ_i

σ_j : Nilai kolom dari total μ_j

c. *Energy*

Merupakan jumlah dari nilai kuadrat pada GLCM yang telah dinormalisasi.

$$Eng = \sum_i \sum_j p_{(i,j)}^2 \quad (3)$$

Keterangan:

Eng : Nilai *energy*.

i : Indeks baris pada matrik.

j : Indeks kolom pada matrik.

d. *Homogeneity*

Merupakan nilai yang berhubungan dengan pengukuran kedekatan antara distribusi nilai elemen yang ada pada GLCM dengan diagonal GLCM.

$$Hom = \sum_i \sum_j \frac{p(i,j)}{1+|i-j|} \quad (4)$$

Keterangan:

Hom : Nilai *homogeneity*.

i : Indeks baris pada matrik.

j : Indeks kolom pada matrik.

2.2 Algoritma Naïve Bayes

Naïve Bayes Classifier merupakan suatu pengklasifikasi yang menggunakan probabilitas bersyarat untuk menghitung probabilitas suatu kejadian/objek masuk ke dalam kategori yang tersedia/diberikan[1]. Dalam *Naïve Bayes Classifier*, diperlukan dua himpunan utama, yaitu himpunan latih dan himpunan uji. Himpunan latih digunakan sebagai data latih yang nantinya sebagai pengidentifikasi objek yang ingin di klasifikasi, sedangkan himpunan uji adalah data yang nantinya akan gunakan sebagai pengukur akurasi dan ketepatan dari pengklasifikasian tersebut. Berikut merupakan teorema *Naïve Bayes*:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (5)$$

Keterangan:

C : Merepresentasikan Kelas

F : Merepresentasikan karakteristik

2.2 Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining[]. *Confusion matrix* biasanya digambarkan dengan sebuah tabel yang didalamnya terdapat jumlah data pengujian yang telah diklasifikasikan antara pengujian benar dan pengujian salah.

Tabel 1. Tabel Confusion Matrix

		Prediksi	
		Positive	Negative
Aktual	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Keterangan :

True Positive (TP) : Data positif yang diprediksi benar

True Negative (TN) : Data negative yang diprediksi benar

False Positive (FP) : Data negative namun diprediksi sebagai data positif

False Negative (FN) : Data positif namun diprediksi sebagai data negative

Keempat parameter tersebut digunakan untuk menghitung

Akurasi

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (6)$$

3 METODOLOGI PENELITIAN

3.1 Akuisisi Data

Akuisisi data dilakukan untuk mendapatkan data sampling yang dibutuhkan dalam penelitian ini. Objek yang digunakan yaitu menggunakan 3 jenis bunga anggrek, yaitu anggrek *dendrobium*, anggrek *Phalaenopsis*, Anggrek *Vanda*. Objek diambil foto dengan menggunakan kamera dan citra dengan *background* putih menggunakan kertas *Art Paper* serta dengan bantuan pencahayaan matahari. Data yang didapat berjumlah 114 citra yang dibagi menjadi 96 data latih dan 18 data uji, setiap jenisnya menggunakan 2 objek bunga dan setiap objek diambil 19 citra yang dimana citra tersebut terdiri dari 8 angle mata angin, 3 arah derajat Bunga tampak depan, 4 arah derajat bunga tampak belakang dan 4 arah derajat bunga tampak atas.

3.2 Praproses Citra

Pada Tahapan ini untuk mendapatkan citra dengan ketentuan yang sama dengan maksud mempermudah sistem dalam melakukan pengolahan citra dilakukan *cropping* dan kemudian data citra yang digunakan diubah menjadi citra *grayscale* dengan tujuan agar dapat menampilkan 1 warna pada setiap citra setelah itu dilakukan *noise removal* untuk memperhalus citra sehingga dalam proses nanti dapat menghasilkan hasil yang maksimal yang kemudian dilakukan pendeteksian garis tepi *canny* yang berguna sebagai segmentasi citra. pada setiap praproses yang dilakukan dengan maksud untuk memberikan hasil yang baik pada saat dilakukan ekstraksi ciri tekstur.

3.3 Ekstraksi Ciri

Pada tahap selanjutnya dilakukan pengambilan nilai-nilai tekstur menggunakan metode GLCM pada citra bunga anggrek agar nilai dari ekstraksi ciri dapat dijadikan sebagai model data latih dan data uji. Proses ekstraksi ciri GLCM yaitu memperhitungkan pasangan dua piksel citra asli, yang dapat diartikan banyaknya kejadian pada satu level piksel yang bertetangga dengan nilai piksel yang lainnya berdasar jarak ketentangan yaitu sebesar 1 piksel dan arah suatu sudut yang kemudian dilakukan pencarian fitur GLCM, sudut yang di gunakan adalah $0^\circ, 45^\circ, 90^\circ, 135^\circ$ yang kemudian dihitung nilai rata-rata terhadap masing-masing nilai fitur yang didapat dari derajat yang digunakan.

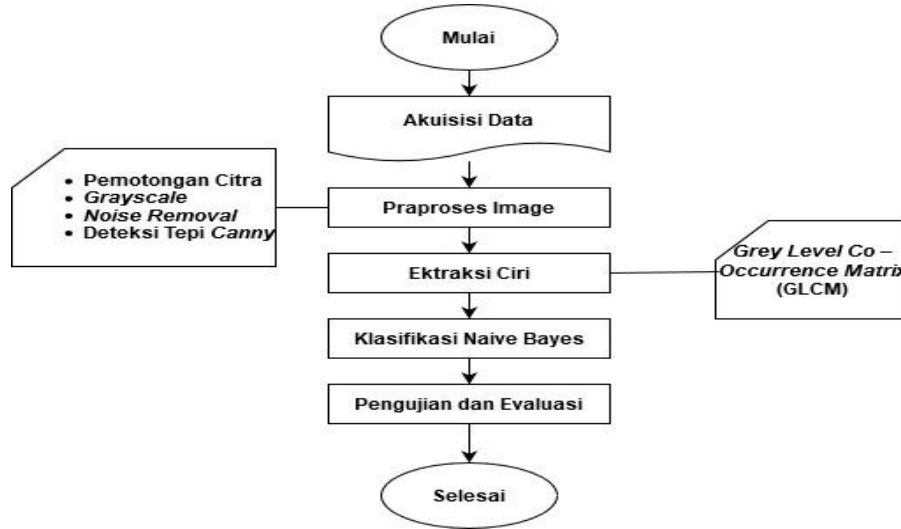
3.4 Klasifikasi Naïve Bayes

Pada tahap ini nilai ekstraksi ciri dari citra akan diklasifikasikan dengan menggunakan algoritma Naïve Bayes, yang dimana nilai ekstraksi ciri data latih akan menjadi sebuah model dataset dan nilai ekstraksi ciri data uji akan digunakan sebagai penguji sistem yang mengacu pada perhitungan probabilitas terhadap data latih. Pengerjaan menggunakan naïve bayes nantinya akan menghasilkan klasifikasi berupa jenis bunga anggrek dendrobium, phalaenopsis dan vanda. Klasifikasi bunga anggrek menggunakan naïve bayes berdasarkan persamaan (5).

3.5 Pengujian dan Evaluasi

Pada tahap ini dilakukan pengujian hasil klasifikasi dengan perhitungan probabilitas kemiripan data lama yang ada dengan data yang baru. Melakukan perbandingan nilai hasil probabilitas menggunakan metode confusion matrix dengan sejumlah data uji. Pada tahapan ini dilakukan perhitungan akurasi.

Alur tahapan penelitian ini akan ditunjukkan pada gambar 1



Gambar. 1. Alur Tahapan Penelitian, berisi setiap proses yang dilakukan dalam pengerjaan pada penelitian ini

4 Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini sebanyak 114 citra bunga angrek yang terbagi menjadi 96 data latih dan 18 data uji. Data citra terdiri dari 3 jenis bunga angrek yang setiap jenisnya menggunakan 2 objek bunga setiap objek diambil 19 citra yang dimana citra tersebut terdiri dari 8 angle mata angin, 3 arah derajat Bunga tampak depan, 4 arah derajat bunga tampak belakang dan 4 arah derajat bunga tampak atas. Ekstraksi ciri menggunakan jarak 1 piksel dan menggunakan derajat 0,45,90,135 yang kemudian dihitung nilai rata-rata dari setiap nilai fitur yang didapat.

Tabel 2. Hasil Ekstraksi Ciri Data Latih

No.	Contrast	Correlation	Energy	Homogeneity	Kelas
1	4,1545	0,293451	0,802898	0,925812	Dendrobium
2	3,9597	0,323482	0,806507	0,929291	Dendrobium
3	3,75642	0,302065	0,819753	0,932921	Dendrobium
4	2,36971	0,324612	0,882528	0,957684	Dendrobium
5	2,40175	0,342958	0,878886	0,957112	Dendrobium
6	2,05829	0,338192	0,896371	0,963245	Dendrobium
7	2,96655	0,349913	0,850138	0,947026	Dendrobium
8	2,51377	0,330789	0,874814	0,955111	Dendrobium

No.	Contrast	Correlation	Energy	Homogeneity	Kelas
9	2,4393	0,325967	0,879029	0,956441	Dendrobium
10	2,33966	0,315118	0,884946	0,95822	Dendrobium
...
...
31	2,86461	0,316904	0,85952	0,948846	Dendrobium
32	2,57819	0,32183	0,872649	0,953961	Dendrobium
33	5,29873	0,291933	0,751142	0,90538	Phalaenopsis
34	7,75508	0,263881	0,652357	0,861516	Phalaenopsis
35	5,11233	0,289998	0,759975	0,908708	Phalaenopsis
36	3,74059	0,319291	0,817507	0,933204	Phalaenopsis
37	4,45752	0,297464	0,788125	0,920401	Phalaenopsis
38	4,898	0,29847	0,767801	0,912536	Phalaenopsis
39	4,1188	0,311566	0,801175	0,92645	Phalaenopsis
40	6,45077	0,290658	0,700475	0,884808	Phalaenopsis
41	5,46809	0,294256	0,743113	0,902355	Phalaenopsis
42	4,56448	0,31382	0,780015	0,918491	Phalaenopsis
...
...
63	3,86533	0,303308	0,814349	0,930976	Phalaenopsis
64	4,58691	0,299703	0,781751	0,781751	Phalaenopsis
65	3,11061	0,330137	0,84588	0,944453	Vanda
66	3,03627	0,338546	0,84831	0,945781	Vanda
67	2,79284	0,326832	0,861756	0,950128	Vanda
68	2,92178	0,348174	0,85256	0,947825	Vanda
69	2,29964	0,339272	0,884329	0,958935	Vanda
70	3,42698	0,317157	0,832673	0,938804	Vanda
71	3,7059	0,322385	0,81865	0,933823	Vanda

No.	Contrast	Correlation	Energy	Homogeneity	Kelas
72	4,07148	0,316503	0,802456	0,927295	Vanda
73	3,38233	0,334317	0,832232	0,939601	Vanda
74	3,01657	0,346711	0,848154	0,946133	Vanda
...
...
95	2,93965	0,308062	0,857093	0,947506	Vanda
96	2,79032	0,302996	0,864762	0,950173	Vanda

Tabel 3. Hasil Ekstraksi Ciri Data Uji

No.	Contrast	Correlation	Energy	Homogeneity	Kelas
1	289.955	0.32463	0.856853	0.948222	Dendrobium
2	360.638	0.326521	0.82271	0.9356	Dendrobium
...
5	251.117	0.319833	0.876152	0.955158	Dendrobium
6	332.504	0.309315	0.838686	0.940624	Dendrobium
7	462.703	0.298182	0.780361	0.917375	Phalaenopsis
8	615.505	0.277657	0.716689	0.890088	Phalaenopsis
...
11	439.525	0.310065	0.788572	0.921513	Phalaenopsis
12	410.898	0.30416	0.802894	0.926625	Phalaenopsis
13	410.646	0.310652	0.801834	0.92667	Vanda
14	187.906	0.324866	0.906388	0.966445	Vanda
...
17	311.883	0.301501	0.301501	0.944307	Vanda
18	445.237	0.305745	0.786757	0.920493	Vanda

Dari hasil ekstraksi ciri diatas yang kemudian akan di proses melalui klasifikasi Naïve Bayes dengan menghitung probabilitas atribut ciri data uji yang mengacu pada data latih. Proses pengujian dilakukan dengan citra yang dimasukkan menghasilkan nilai probabilitas terbesarnya yang kemudian dilakukan pencarian nilai probabilitas terbesar terhadap model yang telah dibentuk menggunakan algoritma *Naïve Bayes Classifier* sebelumnya.

Kelas	Predict			Jumlah Data	Akurasi	
	<i>Dendrobium</i>	<i>Phalaenopsis</i>	<i>Vanda</i>			
Aktual	<i>Dendrobium</i>	4	0	2	6	66,7%
	<i>Phalaenopsis</i>	0	6	0	6	100,0%
	<i>Vanda</i>	2	3	1	6	16,7%
Rata - rata				18	61,1%	

Tabel 4. Evaluasi menggunakan *Confusion Matrix* Data Uji

Nilai akurasi pada penelitian ini sebesar 61,1%, dari hasil pengujian yang didapat, citra bunga anggrek *Vanda* memiliki tingkat kesalahan yang besar sedangkan bunga anggrek *Dendrobium* dan *Phalaenopsis* dapat dikenali walaupun memiliki hasil yang kurang baik. Hal ini dikarenakan nilai ekstraksi ciri bunga anggrek *Vanda* yang mendekati nilai bunga anggrek *Dendrobium* dan *Phalaenopsis*, sehingga algoritma *naïve bayes* sulit dalam melakukan klasifikasi. Hasil dari penerapan algoritma *Naïve Bayes Classifier* dengan ekstraksi ciri GLCM memiliki akurasi yang kurang baik, dikarena hasil dari deteksi *canny* berupa gambar hitam dan putih dan hanya berbentuk sebuah garis dengan nilai dalam piksel 0 dan 255 yang kemudian dilanjut dengan ekstraksi ciri GLCM sehingga proses GLCM menghasilkan nilai yang berdekatan antar kelasnya, sehingga proses pengenalan berjalan kurang baik.

5 Kesimpulan

Berdasarkan hasil penelitian yang diperoleh dapat ditarik kesimpulan Algoritma *Naïve Bayes* dapat melakukan klasifikasi citra bunga anggrek berdasarkan nilai ekstraksi tekstur dengan menggunakan metode GLCM. Karena nilai data yang dihasilkan oleh metode GLCM berupa sebuah data nominal, sedangkan klasifikasi *naïve bayes* dapat menggunakan input berupa data kategorikal maupun data nominal.

Pada evaluasi menggunakan confusion matrik didapat nilai akurasi sebesar 61,1% dikarenakan hasil ekstraksi ciri GLCM yang telah melalui proses deteksi tepi *canny* terlebih dahulu menghasilkan jarak nilai antar kelas yang berdekatan, sehingga proses mengklasifikasi tidak optimal.

Referensi

- [1] Kosasih, Rifki., dan Achmad Fahrurozi. 2017. *Pengklasifikasian Bunga Dengan Menggunakan Metode Isomap dan Naïve Bayes Classifier*. Jurnal Ilmiah Informatika dan Komputer. Vol. 22, No. 3.
- [2] Pamungkas, Damar Putra. 2019. *Ekstraksi Citra Menggunakan Metode GLCM dan KNN untuk Identifikasi Jenis Anggrek (Orchidaceae)*. Innovation In Research Of Informatics. Vol. 1, No. 2.
- [3] Haba, Abdul R. K., dan Kartika Chandra P. 2020. *Pengelompokan Buah Jeruk Menggunakan Naïve Bayes dan Gray Level Co-occurrence Matrix*. ILKOM Jurnal Ilmiah. Vol 12. No 1.
- [4] Yudiyanto, Arief., dan Murinto. 2011. *Pengolahan Citra Digital Untuk Mendeteksi Obyek Menggunakan Pengolahan Warna Model Normalisasi RGB*. Seminar Nasional Teknologi Informasi & Komunikasi Terapan. Vol 2, No. 3.
- [5] Kusumanto, R.D., dkk. 2011. *Klasifikasi Warna Menggunakan Pengolahan Model Warna HSV*. Jurnal Ilmiah Elite Elektro, Vol. 2, No.2.
- [6] Haris, Abdul., dan Andi Prasetyo. 2016. Implementasi Metode Deteksi Tepi Canny Pada Objek Sebagai Model Keamanan Aplikasi Pada Smartphone Android. PETIR. Vol. 9, No. 1.
- [7] Rahmanti, dkk. 2017. *Plasmodium Falciparum Identification in Thick Blood Preparations Using GLCM and Support Vector Machine (SVM)*. Journal Of Applied Intelligent System. Vol. 2, No. 1.