

Klasifikasi Citra Makanan Menggunakan *K-Nearest Neighbor* dengan Fitur Bentuk *Simple Morphological Shape Descriptors* dan Fitur Warna *Grayscale Histogram*

Muhammad Rizky Setiawan¹, Yuita Arum Sari², Putra Pandu Adikara³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹rizky03setiawan@student.ub.ac.id, ²yuita@ub.ac.id, ³adikara.putra@ub.ac.id

Abstrak

Makanan merupakan salah satu sumber energi yang diperlukan oleh manusia. Jenis makanan yang dikonsumsi sangat berpengaruh terhadap daya tahan tubuh. Namun keanekaragaman makanan yang ada menyebabkan masyarakat sulit mengenali jenis makanan yang ingin dikonsumsi. Diperlukannya suatu sistem yang dapat mengenali jenis makanan untuk mempermudah masyarakat dalam mengatur pola makannya. Sebelum memasuki proses ekstraksi fitur, tahap pertama melakukan *preprocessing* dengan memisahkan *background* dengan objek citra makanan. Selanjutnya dilakukan ekstraksi fitur warna menggunakan metode *Grayscale Histogram*. Metode *Grayscale Histogram* menghasilkan fitur *mean*, *standard deviation*, *skewness*. Kemudian dilakukan ekstraksi fitur bentuk menggunakan metode *Simple Morphological Shape Descriptors* (SMSD) dan menghasilkan fitur *area*, *length*, *width*, *aspect ratio*, *rectangularity N*. Setelah didapatkan hasil ekstraksi fitur, dilakukan klasifikasi menggunakan metode *K-Nearest Neighbor*. Berdasarkan hasil pengujian jika hanya menggunakan metode *Grayscale Histogram* menghasilkan nilai akurasi sebesar 60%. Jika hanya menggunakan metode SMSD menghasilkan nilai akurasi sebesar 54,8%. Jika menggunakan metode *Grayscale Histogram* dan metode SMSD menghasilkan nilai akurasi sebesar 77,8%. Metode *Grayscale Histogram* dan metode SMSD dapat digunakan untuk klasifikasi citra makanan menggunakan metode klasifikasi *K-Nearest Neighbor*.

Kata kunci: Makanan, Klasifikasi, *Grayscale Histogram*, *Simple Morphological Shape Descriptors*, *K-Nearest Neighbor*

Abstract

Food is one of the energy sources needed by humans. The type of food consumed greatly affect the immune system. But the diversity of existing food causes people to be difficult to recognize the type of food they want to consume. The need for a system that can recognize types of food to make it easier for people to regulate their diet. Before entering the feature extraction process, the first step is to do preprocessing by separating the background from the food image object. Furthermore, color feature extraction is performed using the Grayscale Histogram method. The Grayscale Histogram method produces the mean, standard deviation, skewness features. Then form feature extraction was performed using the Simple Morphological Shape Descriptors (SMSD) method and produced area features, length, width, aspect ratio, rectangular N. After extracting feature results, classification was done using the K-Nearest Neighbor method. Based on the test results if only using the Grayscale Histogram method produces an accuracy value of 60%. If only using the SMSD method produces an accuracy value of 54.8%. If using the Grayscale Histogram method and the SMSD method produces an accuracy value of 77.8%. The Grayscale Histogram method and the SMSD method can be used to process images using the K-Nearest Neighbor classification method.

Keywords: Food, Classification, *Grayscale Histogram*, *Simple Morphological Shape Descriptors*, *K-Nearest Neighbor*

1. PENDAHULUAN

Makanan merupakan salah satu sumber

energi yang diperlukan oleh manusia (Saparinto & Hidayati, 2010). Daya tahan tubuh dapat dipengaruhi oleh jenis makanan yang dikonsumsi. Masyarakat dapat mengatur pola

makan yang sehat dengan mengetahui jenis makanan. Namun, keanekaragaman makanan yang ada menyebabkan masyarakat sulit mengenali jenis makanan yang ingin dikonsumsi. Pengenalan jenis makanan diperlukan untuk mempermudah dan mempercepat waktu dalam mengidentifikasi makanan.

Suatu proses pengelompokan dua atau lebih data sesuai dengan kesamaan kriteria disebut klasifikasi. Metode klasifikasi yang dapat digunakan untuk mengklasifikasikan citra ada bermacam-macam. Contohnya metode *Naive Bayes* yang digunakan pada klasifikasi citra buah apel menghasilkan akurasi sebesar 63% (Ciputra, et al., 2018). Selain metode *Naive Bayes*, metode lain yang dapat digunakan dalam klasifikasi yaitu *K-Nearest neighbor* (KNN). Metode KNN digunakan dalam mengklasifikasikan daging sapi dengan daging babi dan menghasilkan akurasi sebesar 86,87% (Budianita, 2015).

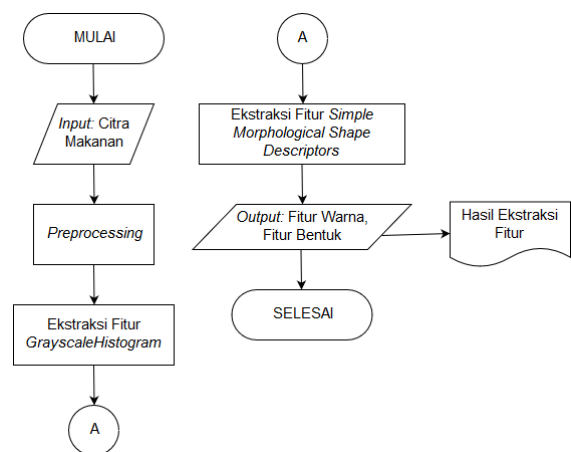
Dalam metode klasifikasi, diperlukan fitur-fitur yang digunakan untuk diklasifikasikan. Salah satu fitur yang digunakan dari citra yaitu fitur bentuk. Fitur bentuk dapat diekstraksi menggunakan metode *Simple Morphological Shape Descriptors* (SMSD). Metode tersebut digunakan pada penelitian sebelumnya mengenai pengenalan spesies tanaman dan menghasilkan akurasi sebesar 87,61% (Wäldchen & Mäder, 2016). Selain fitur bentuk, ada juga fitur lainnya yang bisa digunakan dalam mengenali citra yaitu fitur warna. Fitur warna dapat diekstraksi menggunakan metode *Color Histogram* (Mali & L., 2014). Metode tersebut sangat cepat dan efisien dalam perbandingan komputasinya. Selain itu jika berhubungan dengan warna, kecepatan dalam identifikasi perbedaan citra sangat cepat (Mali & L., 2014). Pada penelitian yang dilakukan oleh Kumar, menjelaskan bahwa channel *grayscale* menghasilkan hasil klasifikasi yang lebih baik dibandingkan menggunakan channel RGB (Kumar, et al., 2016).

Berdasarkan penjelasan yang sudah dijelaskan sebelumnya, metode yang digunakan dalam penelitian ini adalah metode *K-Nearest Neighbor* (KNN) untuk klasifikasi pada citra makanan serta metode *Grayscale Histogram* untuk ekstraksi fitur warna dan metode *Simple Morphological Shape Descriptors* (SMSD) untuk ekstraksi fitur bentuk.

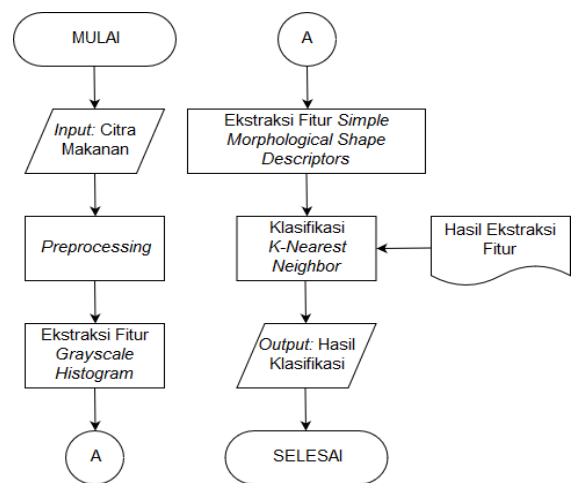
2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini terbagi menjadi dua yaitu pelatihan dan pengujian. Tahap pelatihan dan pengujian diawali dengan input citra makanan yang akan diekstraksi. Kemudian *preprocessing* dilakukan untuk mengambil objek makanan pada gambar. Setelah itu, dilakukan ekstraksi fitur warna *Grayscale Histogram* dan fitur bentuk SMSD. Setelah didapatkan nilai dari setiap fiturnya, pada tahap pelatihan, nilai fitur tersebut disimpan untuk menjadi data latih. Diagram alir untuk metode pelatihan ada pada Gambar 1.

Pada tahap pengujian setelah didapatkan nilai dari setiap fiturnya, dilakukan klasifikasi menggunakan *K-Nearest Neighbor* dan menghasilkan keluaran hasil dari klasifikasi. Diagram alir untuk metode pengujian ada pada Gambar 2.



Gambar 1. Diagram Alir Metode Pelatihan



Gambar 2. Diagram Alir Metode Pengujian

2.1 Simple Morphological Shape Descriptors

(SMSD)

SMSD merupakan metode ekstraksi fitur bentuk dari citra. Ekstraksi fitur bentuk yang dihasilkan dari metode SMSD adalah *area*, *length*, *width*, *aspect ratio*, dan *rectangularity N*. *Area* merupakan jumlah piksel yang ada pada objek. *Length* dan *Width* merupakan panjang dan lebar dari objek pada citra. *Aspect ratio* merupakan rasio dari panjang dan lebar yang ada pada objek. Dan *rectangularity N* merepresentasikan luas minimum persegi panjang yang diisi oleh objek (Wäldchen & Mäder, 2016). Fungsi untuk menghitung nilai *aspect ratio* dan *rectangularity N* ditunjukkan pada Persamaan 1 dan Persamaan 2.

$$rectangularity = \frac{A}{LW} \tag{1}$$

$$aspect\ ratio = \frac{L}{W} \tag{2}$$

Keterangan:

A: Area

L: Major axis length

W: Minor axis length

2.2 Grayscale Histogram

Grayscale Histogram merupakan metode ekstraksi fitur warna dari citra grayscale. Kelebihan dari metode Grayscale Histogram ini adalah dalam komputasinya sangat cepat dan efisien. Ekstraksi fitur warna yang dihasilkan dari metode Grayscale Histogram adalah *mean*, *standard deviation*, dan *skewness*. Sebelum mendapatkan fitur warna, harus dilakukan perhitungan probabilitas *order* pertama histogram yang ditunjukkan pada Persamaan 3.

$$P(g) = \frac{N(g)}{M} \tag{3}$$

Fungsi *mean* ditunjukkan pada Persamaan 4.

$$Mean = \bar{g} = \sum_{g=0}^{L-1} gP(g) \tag{4}$$

Fungsi *standard deviation* ditunjukkan pada Persamaan 5.

$$\sigma_g = \sqrt{\sum_{g=0}^{L-1} (g - \bar{g})^2 P(g)} \tag{5}$$

Fungsi *skewness* ditunjukkan pada Persamaan 6.

$$SKEW = \frac{1}{\sigma_g^3} \sum_{g=0}^{L-1} (g - \bar{g})^3 P(g) \tag{6}$$

Keterangan:

P(g): probabilitas gray level g pada citra

N(g): nilai piksel gray level g pada citra

M: total piksel pada citra

g: gray level

σ : standard deviation

\bar{g} : mean

2.3 K-Nearest Neighbor (KNN)

KNN merupakan suatu metode klasifikasi yang sederhana dan efektif dan melakukan klasifikasi sesuai dengan banyaknya kedekatan jarak yang mayoritas terhadap data yang diklasifikasi (Santoso, 2007). Metode KNN dapat menghasilkan akurasi yang akurat ketika data *training* yang digunakan banyak. Metode KNN melakukan klasifikasi pada data berdasarkan data latih yang sudah diklasifikasikan.

Sebelum menghitung jarak antar data, pertama harus menentukan nilai *k* yang akan digunakan. Kemudian menghitung jarak *Euclidean* data yang ingin diklasifikasikan dengan data latih menggunakan Persamaan 7.

$$D(a, b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2} \tag{7}$$

Keterangan:

D(a,b) = jarak antara vektor a dengan vektor b

a: vektor a

b: vektor b

k: jumlah fitur yang digunakan

Setelah didapatkan jaraknya, maka jarak diurutkan mulai dari yang terkecil sampai terbesar. Kemudian mengklasifikasikan data sejumlah *k* dan melihat kelas mayoritas yang ada. Hasil dari kelas mayoritas menjadi kelas data yang baru.

3. HASIL DAN PEMBAHASAN

Pengujian yang dilakukan adalah pengujian terhadap rasio jumlah data, pengujian terhadap nilai *k* pada metode klasifikasi KNN, dan pengujian terhadap masing-masing fitur warna dan bentuk. Kelas yang digunakan berjumlah 12 yaitu donat (001), mie (004), ayam krispi (008), daging rendang (009), selada (013), tomat (015), pisang hijau (017), pisang kuning (018), nasi kuning (021), happy toss (027), biskuit (029), genie pie (032).

3.1. Pengujian Rasio Jumlah Data

Pengujian rasio jumlah data dilakukan untuk mengetahui hubungan antara jumlah data yang digunakan dengan akurasi yang dihasilkan. Hasil pengujian terhadap rasio jumlah data ditunjukkan pada Tabel 1.

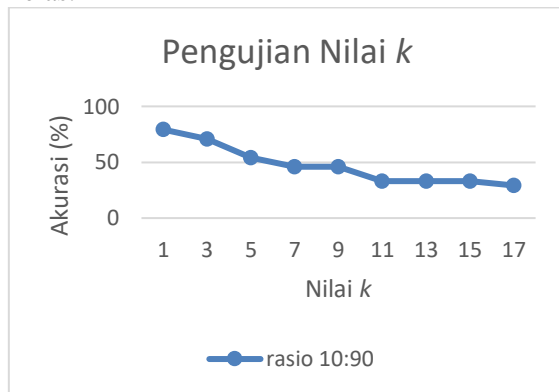
Pengujian rasio jumlah data menghasilkan

akurasi tertinggi yaitu 70,8% dengan perbandingan rasio 10:90 dan akurasi terendah yaitu 58,3% dengan perbandingan rasio 30:70. Penurunan nilai akurasi ini disebabkan oleh berkurangnya rasio data *testing* dan *training* yang digunakan. Semakin besar rasio yang digunakan maka hasil dari klasifikasinya semakin akurat.

3.2. Pengujian Nilai *k*

Hasil pengujian nilai *k* pada metode klasifikasi KNN ditunjukkan pada Tabel 2. Gambar 3 menunjukkan grafik perbedaan hasil akurasi dari pengujian nilai *k*.

Berdasarkan Gambar 3, pengujian nilai *k* menghasilkan akurasi tertinggi yaitu 79,2% dengan nilai *k*=1 dan akurasi terendah yaitu 29,2% dengan nilai *k*=17. Penurunan nilai akurasi ini disebabkan karena jumlah kelas ketetanggaan semakin bertambah ketika nilai *k* semakin mendekati jumlah data pada setiap kelas.



Gambar 3. Grafik Hasil Pengujian Nilai *k*

Tabel 1. Hasil Akurasi Pengujian Rasio Jumlah Data

Nilai <i>k</i>	Perbandingan Data Testing:Training		
	10:90	20:80	30:70
3	70,8%	64,6%	58,3%

Tabel 2. Hasil Akurasi Pengujian Nilai *k*

Nilai <i>k</i>	Rasio Jumlah Data 10:90
1	79,2%
3	70,8%
5	54,2%
7	45,8%
9	45,8%
11	33,3%
13	33,3%
15	33,3%
17	29,2%

3.3. Pengujian Fitur Warna *Grayscale Histogram* dan Fitur Bentuk SMSD

Pengujian fitur dibagi menjadi tiga, pertama menggunakan fitur warna *Grayscale Histogram* saja, kedua menggunakan fitur bentuk SMSD saja, dan ketiga menggunakan fitur warna *Grayscale Histogram* dan fitur bentuk SMSD. Hasil dari pengujian ketiganya ditunjukkan pada Tabel 3, Tabel 4, dan Tabel 5. Contoh hasil klasifikasi menggunakan fitur warna dan fitur bentuk ditunjukkan pada Tabel 6.

Tabel 3. Hasil Pengujian Fitur *Grayscale Histogram*

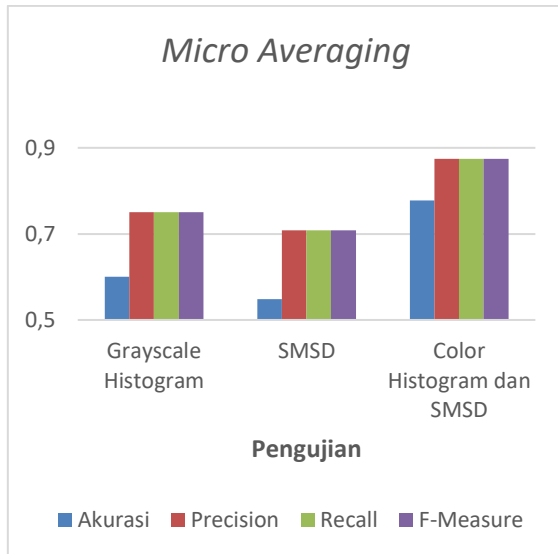
No	Kelas	Akurasi	Precision	Recall	F-Measure
1	001	1	1	1	1
2	004	1	1	1	1
3	008	1	1	1	1
4	009	0,5	1	0,5	0,667
5	013	1	1	1	1
6	015	0,333	0,5	0,5	0,5
7	017	0,667	0,667	1	0,8
8	018	0,667	0,667	1	0,8
9	021	1	1	1	1
10	027	0,667	0,667	1	0,8
11	029	0	0	0	0
12	032	0	0	0	0

Tabel 4. Hasil Pengujian Fitur Bentuk SMSD

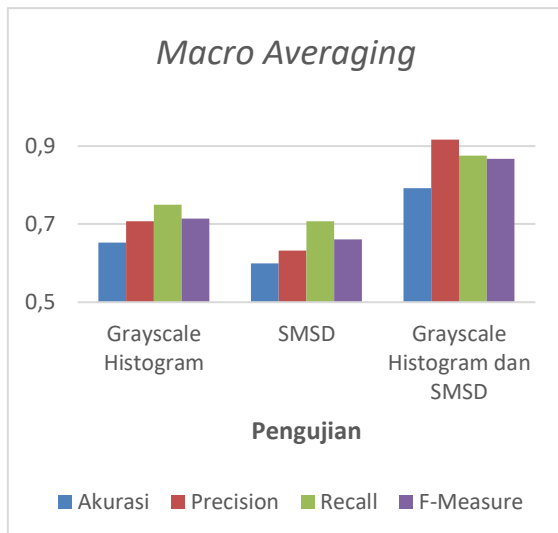
No	Kelas	Akurasi	Precision	Recall	F-Measure
1	001	1	1	1	1
2	004	0,333	0,5	0,5	0,5
3	008	0,2	0,25	0,5	0,333
4	009	1	1	1	1
5	013	1	1	1	1
6	015	0	0	0	0
7	017	1	1	1	1
8	018	0,667	0,667	1	0,8
9	021	0,667	0,667	1	0,8
10	027	0,333	0,5	0,5	0,5
11	029	1	1	1	1
12	032	0	0	0	0

Tabel 5. Hasil Pengujian Fitur Warna *Grayscale Histogram* dan Fitur Bentuk SMSD

No	Kelas	Akurasi	Precision	Recall	F-Measure
1	001	1	1	1	1
2	004	0,5	1	0,5	0,667
3	008	0,667	0,667	1	0,8
4	009	1	1	1	1
5	013	1	1	1	1
6	015	0,5	1	0,5	0,667
7	017	1	1	1	1
8	018	0,667	0,667	1	0,8
9	021	0,667	0,667	1	0,8
10	027	1	1	1	1
11	029	1	1	1	1
12	032	0,5	1	0,5	0,667







Gambar 4. Grafik Perbandingan Pengujian Fitur menggunakan *Micro-Averaging*



Gambar 5. Grafik Perbandingan Pengujian Fitur menggunakan *Macro-Averaging*

Tabel 6. Contoh Hasil Klasifikasi Menggunakan Fitur Warna *Grayscale Histogram* dan Fitur Bentuk SMSD

No	Data Uji	Hasil Sebenarnya	Hasil Klasifikasi
1		004	004
2		015	015
3		015	027
4		004	008

Berdasarkan Tabel 2, Tabel 3, dan Tabel 4, dihitung nilai akurasi, *precision*, *recall*, dan *f-measure* menggunakan *micro-averaging* dan *macro-averaging*. Dari hasil yang ditunjukkan pada Gambar 4, dihasilkan nilai akurasi, *precision*, *recall* dan *f-measure* dengan menggunakan *micro-averaging*.

Nilai tertinggi akurasi, *precision*, *recall*, dan *f-measure* dari ketiga pengujian didapatkan ketika menggunakan kombinasi antara fitur warna *Grayscale Histogram* dengan fitur Bentuk SMSD dengan nilai 0,778, 0,875, 0,875, dan 0,875. Nilai dari akurasi, *precision*, *recall*, dan *f-measure* dengan menggunakan *macro-averaging* yang ditunjukkan pada Gambar 5 tidak seimbang karena nilai dibagi dengan masing-masing kelas yang digunakan. Nilai tertinggi akurasi, *precision*, *recall*, dan *f-measure* dari ketiga pengujian didapatkan ketika menggunakan kombinasi antara fitur warna *Grayscale Histogram* dengan fitur Bentuk SMSD dengan nilai 0,792, 0,917, 0,875, dan 0,867. Pada pengujian fitur SMSD menghasilkan nilai yang lebih rendah dibandingkan dengan hasil dari pengujian fitur *Grayscale Histogram*. Namun, jarak dari nilai fitur SMSD dengan fitur *Grayscale Histogram* tidak begitu jauh yaitu 0,052 untuk nilai akurasinya. Hal ini membuktikan bahwa kedua fitur tidak saling mendominasi ketika dilakukan klasifikasi.

4. KESIMPULAN

Metode ekstraksi fitur *Grayscale Histogram* dan fitur SMSD dapat digunakan dalam klasifikasi menggunakan KNN dalam melakukan klasifikasi citra makanan. Nilai *k* yang menghasilkan nilai akurasi terbaik dengan menggunakan rasio jumlah data 10:90 yaitu ketika *k=1* dengan nilai akurasi sebesar 77,8% jika menggunakan fitur warna *Grayscale Histogram* dan fitur bentuk SMSD. Jika menggunakan fitur warna *Grayscale Histogram* saja maka nilai akurasi yang dihasilkan sebesar 60%. Jika menggunakan fitur bentuk SMSD dihasilkan nilai akurasi sebesar 54,8%.

Untuk penelitian selanjutnya, diperlukan penambahan data *training* yang digunakan ketika menggunakan metode klasifikasi KNN. Selain itu, diperlukan penambahan ekstraksi fitur agar jumlah fitur pada citra bertambah.

DAFTAR RESENSI

Budianita, E., 2015. Implementasi Pengolahan Citra dan Klasifikasi K-Nearest Neighbour

- Untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi. *Jurnal Sains, Teknologi dan Industri*, Volume 12, pp. 242-247.
- Ciputra, A., Moses Sediadi, D. R. I., Rahmawanto, E. H. & Susanto, A., 2018. KLASIFIKASI TINGKAT KEMATANGAN BUAH APEL MANALAGI DENGAN ALGORITMA NAIVE BAYES DAN EKSTRAKSI FITUR CITRA DIGITAL. *SIMETRIS*, 9(1).
- C, S. A., M, T. & P, A., 2014. A Study on Curvature Scale Space. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(3), pp. 168-173.
- Gonzalez, R. C., 2002. *Digital Image Processing*. 2nd ed. Upper Saddle River New Jersey 07458: Prentice Hall.
- Han, J., Kamber, M. & Pei, J., 2006. *Data Mining Concepts & Techniques*. 2nd ed. San Fransisco: Elsevier.
- Karmilasari & Sumarna, A., 2011. TEMU KEMBALI CITRA BERBASIS CITRA WARNA. *Seminar Nasional Aplikasi Teknologi Informasi*.
- Kumar, K., Li, J.-P., Z.-u.-a. & Shaikh, R. A., 2016. Content Based Image Retrieval Using Gray Scale Weighted Average Method. *International Journal of Advanced Computer Science and Applications*, 7(1).
- Mali, S. N. & L., T. M., 2014. Color Histogram Features for Image Retrieval Systems. *International Journal of Innovative Research in Science, Engineering and Technology*, 3(4).
- Mudjiyanto, T. T., Susanto, D. & Luciasari, E., 1994. KEBIASAAN MAKAN GOLONGAN REMAJA DI ENAM KOTA BESAR DI INDONESIA. In: *Penelitian Gizi dan Makanan*. s.l.:National Institute of Health Research and Development, Ministry of Health of Republic of Indonesia.
- Santoso, B., 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. 1st ed. Yogyakarta: Graha Ilmu.
- Saparinto, C. & Hidayati, 2010. *Bahan Tambahan Pangan*. Yogyakarta: Kanisius.
- Sokolova, M. & Lapalme, G., 2009. A Systematic Analysis of Performance Measures For Classification Tasks. *Information Processing and Management*, 45(4), pp. 427-437.
- Wäldchen, J. & Mäder, P., 2016. Plant Species Identification Using Computer Vision Techiques: A Systematic Literature Review. *Computational Methods In Engineering*.