

Seleksi Fitur *Information Gain* Pada Klasifikasi Citra Makanan Menggunakan *Hue Saturation Value* dan *Gray Level Co-Occurrence Matrix*

Friska Yessy Nabella¹, Yuita Arum Sari², Randy Cahya Wihandika³

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

Abstrak

Pemilihan makanan menjadi masalah terutama bagi mereka yang baru pertama kali mencicipi atau mencari suatu makanan. Klasifikasi penting dilakukan untuk seseorang yang sedang menderita penyakit tertentu, penerapan diet dan untuk menjalani gaya hidup sehat. Pemilihan makanan juga diperlukan bagi para penyandang tunanetra untuk mengetahui nama suatu makanan. Klasifikasi pada penelitian ini diawali dengan *pre-processing* citra yang menghasilkan citra tersegmentasi. Kemudian ekstraksi fitur menggunakan ekstraksi warna *Hue Saturation Value* (HSV) dan tekstur *Gray Level Co-Occurrence Matrix* (GLCM). Berdasarkan fitur yang telah diekstraksi kemudian dilakukan seleksi fitur dengan untuk mengambil fitur-fitur relevan guna meningkatkan akurasi dan mengurangi beban kerja sistem menggunakan *Information Gain*. Proses yang terakhir yaitu klasifikasi menggunakan *K-Nearest Neighbor*. Hasil akurasi terbaik sebesar 95,24% pada fitur HSV saja dengan menggunakan seleksi fitur dengan nilai $k=1$. Pada kombinasi HSV dan GLCM menggunakan *Information Gain* mampu meningkatkan akurasi dari 57,14% menjadi 87,61%. Begitu juga dengan fitur GLCM saja menggunakan *Information Gain* yang dapat meningkatkan akurasi dari 57,14% menjadi 74,28%. Dengan demikian seleksi fitur *Information Gain* sangat berperan dalam meningkatkan akurasi dan mampu mendapatkan fitur-fitur relevan apabila fitur berjumlah banyak. Sedangkan apabila jumlah fitur sedikit, peningkatan akurasi tidak terlalu signifikan namun tetap mampu mengurangi beban kerja sistem.

Kata kunci: klasifikasi, makanan, HSV, GLCM, information gain

Abstract

Choosing food has become a challenge for those who are presented with new food choices. Classification is important for those who have a strict diet regarding food that they consume. Food selection is essential for those who are visually impaired to identify food items. The classification process in this research is initiated with the *pre-processing* of the image, resulting in a segmented image which is then continued with feature extraction where *Hue Saturation Value* (HSV) for color extraction and *Gray Level Co-Occurrence Matrix* (GLCM) for texture features. Based on features that have been extracted the next step is to gather relevant features using *Information gain* to reduce the workload. The last process is classification, using *K-Nearest Neighbor*. Accuracy results are 95,24% at best using only HSV with $k=1$ for feature selection. A combination of HSV and GLCM using *Information gain* results in a accuracy from 57,14% to 87,61%. This also applies to only using GLCM with *information gain* that raises the accuracy from 57,14% to 74,28%. With the previous statement taken into consideration, *Information Gain* as a feature selection method increases accuracy with a significant amount and is able to obtain relevant feature if the list of features is abundant. If there are only a few features used, the accuracy increment is not that significant but it decreases the workload of the system.

Keywords: classification, food, HSV, GLCM, information gain

1. PENDAHULUAN

Memilih makanan merupakan hal penting bagi penderita penyakit tertentu. Contohnya pada penderita jantung *coroner* yang

harus memperhatikan pemenuhan gizi yang tepat agar tidak memicu keadaan jantung menjadi lebih buruk (Puspitasari, Ratnawati, & Fauzi, 2018). Dalam penerapan diet juga dibutuhkan pengaturan menu makan yang tepat untuk

menghindari dan membatasi makanan yang menyebabkan peningkatan kolesterol dan tekanan darah (Purwati, 2002). Para penyandang tunanetra juga kesulitan dalam pemilihan makanan sehingga membutuhkan suatu sistem untuk mengetahui nama suatu makanan (Hoonlor, et al., 2015).

Klasifikasi makanan penting dilakukan untuk menghindari penyakit, menjalani gaya hidup sehat, penerapan diet pada penderita diabetes, seorang penderita alergi makanan serta membantu dalam menemukan kalori makanan, nilai nutrisi dan sebagai referensi makanan. Makanan yang sangat bermacam-macam menjadi tantangan dalam pengklasifikasian makanan. Klasifikasi makanan dapat dilakukan menggunakan metode-metode yang ada pada *computer vision* (Jahan, Kekha, & Quadri, 2018).

Penelitian He, et al. (2014) telah mencoba mengatasi permasalahan tersebut dengan melakukan klasifikasi 42 kelas makanan. Menggunakan *Dominant Color Descriptor* (DCD), *Multi-scale Dense SIFT* (MDSIFT), *Scalable Color Descriptor* (SCD), *Entropy-Based Categorization and Fractal Dimension* (EFD), *Gabor-Based Image Decomposition and Fractal Dimension Estimation* (GFD) dan *Scale Invariant Feature Transforms* (SIFT) dan metode klasifikasi KNN. Hasil dari penelitian ini didapatkan bahwa kombinasi dari fitur DCD, MDSIFT dan SCD menghasilkan akurasi terbaik yaitu 85,1% dimana lebih baik 22% dibandingkan kombinasi fitur lainnya.

Penelitian Mohanaiah, Sathyanarayana, dan Gurukumar (2013) mengatakan terdapat beberapa cara melakukan ekstraksi fitur seperti melalui warna, tekstur, dan bentuk. GLCM merupakan salah satu algoritme ekstraksi tekstur populer yang telah terbukti baik untuk mengekstraksi tekstur dan mendapatkan hasil akurasi yang tinggi. Penelitian lain Hastuti, Dewi, & Widodo (2017) menggunakan GLCM dengan sudut 0°, 45°, 90°, dan 135° menggunakan *K-NN* mendapatkan hasil klasifikasi terbaik menggunakan sudut 0° yaitu dengan akurasi 100%.

Penelitian Alhassan & Khader (2014) menggunakan fitur warna yang diekstraksi oleh *Color Moment* pada ruang warna HSV menunjukkan bahwa penggunaan fitur warna dan tekstur mendapatkan hasil yang lebih akurat daripada hanya menggunakan salah satunya saja. Penelitian (Kusumanto, Pambudi, & Tomponu, 2011) menyatakan bahwa penggunaan ruang

warna HSV dapat mendeteksi dan mengurangi pengaruh intensitas cahaya dari luar. Penelitian (Singh & Hemachandran, 2012) menyebutkan bahwa RGB benar-benar menggambarkan warna yang terlihat sehingga tidak menyediakan titik awal yang berguna untuk mewakili warna gambar dan juga ruang warna RGB tidak memiliki perseptual yang seragam. Maka dibutuhkan untuk menghasilkan ruang warna lain contohnya HSV, CIELAB dan CIELUV. Hasil terbaik dari ketiga ruang warna tersebut yaitu HSV karena memiliki keseragaman perseptual. Ruang warna HSV banyak digunakan dalam bidang penglihatan warna dimana setiap channelnya berhubungan erat dengan kategori persepsi warna manusia.

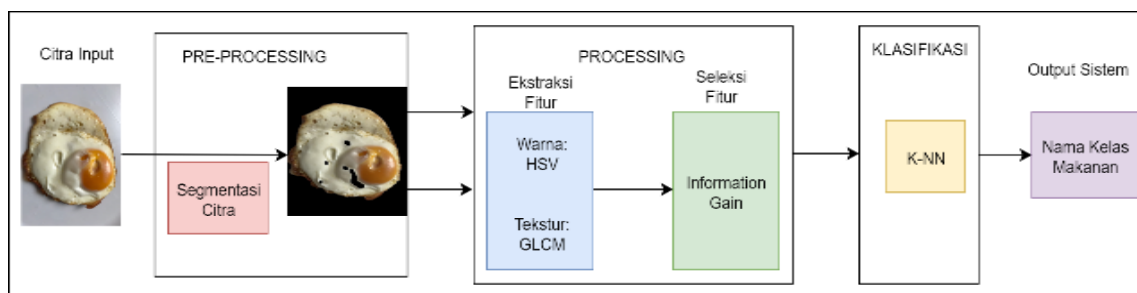
Hasil dari ekstraksi fitur sering kali mendapatkan fitur terlalu banyak yang menyebabkan beban kerja tinggi sehingga menurunkan akurasi. Dengan demikian perlu menghilangkan fitur-fitur tersebut untuk mendapatkan fitur yang relevan dan tidak berlebihan yang dapat memberikan solusi optimal tanpa mengurangi akurasi (Mwadulo, 2016). *Information gain* merupakan metode seleksi fitur yang bekerja dengan cara melakukan pemeringkatan secara sederhana (Chormunge & Jena, 2016). Penelitian oleh Shaltout, et al. (2014) menyebutkan bahwa *information gain* mampu mendeteksi fitur-fitur yang paling relevan dengan menghitung nilai *entropy* berdasarkan kelas-kelas tertentu.

Penelitian Redjeki (2013) membandingkan akurasi dari dua metode yaitu KNN dan *Backpropagation*. Hasil penelitian tersebut menyimpulkan bahwa algoritme KNN mendapatkan hasil identifikasi yang lebih baik daripada algoritme *Backpropagation*. Penelitian lain oleh Priambodo, et al. (2015) melakukan klasifikasi penyakit tanaman jeruk keprok menggunakan metode KNN dengan hasil akurasi 96,67% membuktikan bahwa metode KNN baik dalam mengklasifikasikan citra.

Pada penelitian ini penulis mengusulkan melakukan klasifikasi citra makanan menggunakan KNN dengan penggunaan *Information Gain* pada HSV dan GLCM untuk mendapatkan fitur-fitur yang relevan agar dapat meningkatkan akurasi dan meringankan beban kerja sistem.

2. METODOLOGI PENELITIAN

Tahap awal penelitian ini yaitu melakukan *pre-processing* citra pada data latih dan data uji.



Gambar 1. Strategi Penelitian

Data yang telah dilakukan *pre-processing* akan menghasilkan citra makanan tersegmentasi. Citra tersegmentasi kemudian dilakukan ekstraksi fitur menggunakan HSV dan GLCM. Fitur yang dihasilkan pada *Color Moment* HSV berjumlah 12 fitur yang berasal dari *mean*, *deviasi standar*, *skewness* dan *kurtosis* pada setiap channel HSV. Fitur yang dihasilkan pada GLCM berjumlah 40 fitur. Fitur tersebut berasal dari *mean X*, *mean Y*, *variance X*, *variance Y*, *energy*, *entropy*, *contrast*, *dissimilarity*, *homogeneity* dan *correlation* pada sudut sudut 0°, 45°, 90°, dan 135°. Berdasarkan 52 fitur yang telah didapatkan selanjutnya dilakukan seleksi fitur menggunakan *Information Gain* untuk mendapatkan fitur-fitur relevan sehingga dapat dilakukan pengurangan jumlah fitur pada proses klasifikasi. Setelah mendapatkan fitur-fitur relevan langkah selanjutnya yaitu melakukan klasifikasi menggunakan KNN. Strategi penelitian dalam penelitian ini dapat dilihat pada Gambar 1.

2.1. Data

Data yang digunakan pada penelitian ini merupakan data primer. Data berupa citra makanan berjumlah 21 jenis makanan yang terdiri dari sepuluh data latih dan lima data uji pada setiap kelasnya. Data citra makanan ini diambil di ruang Grup Riset *Computer Vision* gedung F lantai sembilan, Fakultas Ilmu Komputer, Universitas Brawijaya pada tanggal 28 Agustus 2018 pukul 10.00 sampai 13.00 WIB. Contoh data yang digunakan dapat dilihat pada Gambar 2. Data yang diperoleh dari ekstraksi fitur merupakan data numerik, sehingga perlu dilakukan konversi terlebih dahulu menjadi data kategori sebelum masuk ke tahap *Information Gain*. Konversi data numerik menjadi kategori direpresentasikan pada Persamaan satu sampai tiga. Darwiyo, Binawan, & Junaedi (2017)

1. Mencari Jangkauan (J)

$$J = \text{nilai maks} - \text{nilai min} \quad (1)$$

2. Mencari banyak kelas (K)

$$K = 1 + 3,3 \times \log n \quad (2)$$

dimana n adalah jumlah data

3. Menghitung panjang kelas (C)

$$C = \frac{J}{K} \quad (3)$$

4. Memasukkan data ke kategori

2.2. Pre-processing

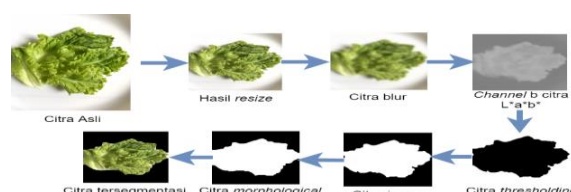
Pre-processing pada penelitian ini bertujuan untuk mendapatkan citra tersegmentasi. Langkah-langkah untuk mendapatkan citra tersegmentasi tersebut ditunjukkan pada Gambar 3.

2.3. Hue Saturation Value (HSV)

Hue yang berarti ukuran panjang gelombang pada warna yang dominan berdasarkan persepsi mata manusia. *Saturation* menggambarkan banyaknya cahaya putih yang bercampur pada *channel hue* (Kusumanto, Pambudi, & Tomponu, 2011). *Brightness* atau *value* adalah intensitas pemantulan objek yang diterima oleh mata. Intensitas tersebut berarti adanya perubahan warna putih menjadi abu-abu sampai hitam. Berikut merupakan rumus dan langkah-langkah untuk mengkonversi RGB ke HSV (OpenCV, 2017).

Kode Kelas	Data Training	Data Testing
0		
3		
7		

Gambar 2. Sampel Dataset

Gambar 3. Langkah-langkah *pre-processing*

1. Normalisasi citra, dengan membagi nilainya dengan 255 yang ditampilkan pada Persamaan 4 sampai 6.

$$H = \frac{H}{255} \quad (4)$$

$$S = \frac{S}{255} \quad (5)$$

$$V = \frac{V}{255} \quad (6)$$

2. Konversi ke HSV, dengan menggunakan Persamaan 7 sampai 9.

$$V = \max(R, G, B) \quad (7)$$

$$S = \begin{cases} \frac{V - \min(R, G, B)}{V} & \text{if } v \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

$$H = \begin{cases} \frac{60(G-B)}{(V - \min(R, G, B))} & \text{if } V = R \\ \frac{120 + 60(B-R)}{(V - \min(R, G, B))} & \text{if } V = G \\ \frac{240 + 60(R-G)}{(V - \min(R, G, B))} & \text{if } V = B \end{cases} \quad (9)$$

3. Mengubah citra ke 8 bit image, dengan menggunakan persamaan 10 sampai 12.

$$V = V \times 255 \quad (10)$$

$$S = S \times 255 \quad (11)$$

$$H = \frac{H}{2} \quad (12)$$

Keterangan:

H: nilai piksel HSV pada *channel* H

S: nilai piksel HSV pada *channel* S

V: nilai piksel HSV pada *channel* V

Fitur Statistik Warna oleh Kadir & Susanto (2013) dalam bukunya, dapat dihitung menggunakan Persamaan 13 sampai 16.

1. Mean

Mean merupakan nilai rata-rata, sehingga memberitahu sesuatu tentang kecerahan umum gambar.

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \quad (13)$$

2. Deviasi Standar

Deviasi standar menjelaskan sesuatu tentang kontras yaitu menggambarkan penyebaran dalam data, yang didefinisikan pada rumus berikut:

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^2} \quad (14)$$

3. Skewness

Skewness atau kecondongan menjelaskan ukuran tentang ketidaksimetrisan, yang didefinisikan pada rumus berikut:

z	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Gambar 4. Pembentukan matriks CM

$$\theta = \frac{\sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^3}{MN\sigma^3} \quad (15)$$

4. Kurtosis

Kurtosis menunjukkan sebaran data bersifat meruncing atau mengumpul, yang didefinisikan pada rumus berikut:

$$\gamma = \frac{\sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^4}{MN\sigma^4} - 3 \quad (16)$$

Keterangan:

M: Panjang citra

N: Lebar citra

P_{ij} : Nilai piksel pada indeks i, j

2.4. Gray Level Co-Occurrence Matrix (GLCM)

Penelitian ini menggunakan GLCM dengan empat sudut dan jarak ketetanggaan satu. Langkah awal yang dilakukan untuk mendapatkan fitur tekstur yaitu melakukan konversi citra RGB menjadi *grayscale*, selanjutnya menentukan ukuran matriks *co-occurrence* (CM). Ukuran matriks *co-occurrence* ditentukan berdasarkan Persamaan 17. Matriks *co-occurrence* dibuat dengan menghitung probabilitas kemunculan nilai ketetanggaan berdasarkan sudutnya (Madenda, 2015). Contoh untuk mendapatkan matriks *co-occurrence* sudut 0 ditunjukkan pada Gambar 4.

$$G = \text{maks} - \text{min} + 1 \quad (17)$$

Keterangan:

maks: nilai maksimum dari citra *grayscale*

maks: nilai minimum dari citra *grayscale*

Setelah mendapatkan matriks *co-occurrence* selanjutnya yaitu melakukan normalisasi untuk menghilangkan ketergantungan pada ukuran citra. Elemen pada GLCM dinormalisasi untuk mendapatkan jumlah senilai satu. Langkah terakhir merupakan perhitungan fitur statistik GLCM berdasarkan elemen GLCM yang telah di normalisasi. Contoh normalisasi matriks *co-occurrence* dapat dilihat pada Gambar 5.

Untuk mendapatkan fitur-fitur statistik GLCM dapat dilakukan dengan melakukan perhitungan berdasarkan rumus-rumus pada persamaan 18 sampai 27 (Madenda, 2015).

$\frac{2}{12}$	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{0}{12}$
$\frac{0}{12}$	$\frac{2}{12}$	$\frac{0}{12}$	$\frac{0}{12}$
$\frac{0}{12}$	$\frac{0}{12}$	$\frac{3}{12}$	$\frac{1}{12}$
$\frac{0}{12}$	$\frac{0}{12}$	$\frac{0}{12}$	$\frac{1}{12}$

Gambar 5. Normalisasi matriks CM

1. Mean

Merupakan rata-rata distribusi probabilitas $P(i, j)$ yang dapat dihitung berdasarkan sampel x dan y , yang didefinisikan pada rumus berikut:

$$\mu_x = \sum_{i=0}^{G-1} i \sum_{j=0}^{G-1} P(i, j) \quad (18)$$

$$\mu_y = \sum_{j=0}^{G-1} j \sum_{i=0}^{G-1} P(i, j) \quad (19)$$

2. Variance

Menentukan sebaran nilai atau simpangan terhadap rata-rata data, yang didefinisikan pada rumus berikut:

$$\sigma_{x^2} = \sum_{i=0}^{G-1} (i - \mu_x)^2 \sum_{j=0}^{G-1} P(i, j) \quad (20)$$

$$\sigma_{y^2} = \sum_{j=0}^{G-1} (j - \mu_y)^2 \sum_{i=0}^{G-1} P(i, j) \quad (21)$$

3. Energy

Energy menunjukkan tingkat keseragaman tekstur, yang didefinisikan pada rumus berikut:

$$Energy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (P(i, j))^2 \quad (22)$$

4. Entropy

Entropy menunjukkan ukuran tingkat keacakan permukaan tekstur akibat gangguan frekuensi, yang didefinisikan pada rumus berikut:

$$Entropy = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \log(P(i, j)) \quad (23)$$

5. Contrast

Contrast menunjukkan nilai intensitas lokal dalam matriks *co-occurrence*, yang didefinisikan pada rumus berikut:

$$Contrast = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - j)^2 P(i, j) \quad (24)$$

6. Dissimilarity

Ukuran yang mendefinisikan variasi tingkat intensitas pasangan piksel dalam citra, yang didefinisikan pada rumus berikut:

$$Dissimilarity = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |i - j| P(i, j) \quad (25)$$

7. Homogeneity

Ukuran perulangan struktur tekstur yang bobot nilainya merupakan nilai *invers* dari *contrast*-nya, yang didefinisikan pada rumus berikut:

$$Homogeneity = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{P(i, j)}{1 + |i - j|} \quad (26)$$

8. Correlation

Correlation mengukur ketergantungan linier dari tingkat abu-abu dari piksel tetangga pada posisi tertentu, yang didefinisikan pada rumus berikut:

$$Correlation = \frac{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{(i - \mu_x)(j - \mu_y)P(i, j)}{\sigma_x \sigma_y}}{\sigma_x \sigma_y} \quad (27)$$

Keterangan:

G: Ukuran citra

i : Array indeks ke i

j : Array indeks ke j

$P(i, j)$: Nilai piksel pada indeks i, j

2.5. Information Gain

Information Gain adalah suatu metode seleksi fitur yang bekerja dengan cara melakukan perangkingan secara sederhana namun telah banyak digunakan dalam berbagai aplikasi salah satunya yaitu pada analisis data citra (Chormunge & Jena, 2016). Information Gain mendeteksi fitur-fitur yang paling relevan berdasarkan kelas-kelas tertentu dengan melakukan perhitungan nilai *entropy*. Entropy adalah ukuran dari ketidakpastian kelas menggunakan probabilitas dari suatu atribut tertentu (Shaltout, et al., 2014). Tahapan dalam proses perhitungan Information Gain dapat dihitung menggunakan persamaan 28 sampai 29. (Firmansyah & Gantini, 2016).

1. Mencari nilai entropi seluruh kelas dengan persamaan 28.

$$Entropy(S) = - \sum_i^c P_i \log_2 P_i \quad (28)$$

2. Mencari nilai gain setiap fitur A dengan rumus Persamaan 29.

$$Gain(S, A) = Entropy(S) - \sum_{Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (29)$$

Keterangan

c : jumlah kelas

P_i : jumlah sampel setiap kelas i

A: fitur

v : nilai untuk fitur A

Values(A) : nilai-nilai untuk fitur A

ISv l: jumlah sampel v
 l S l: jumlah semua data
 Entropy(Sv): *entropy* untuk data yang memiliki nilai v

2.6. K-Nearest Neighbor (KNN)

KNN merupakan salah satu metode klasifikasi yang menggunakan jarak terdekat sebagai pendekatannya. Teknik untuk mencari tetangga terdekat yang dapat dilakukan dengan menggunakan jarak *Euclidean*.

Jarak Euclidean adalah cara mencari jarak antara dua titik dalam ruang dua dimensi (Hastuti, Dewi, & Widodo, 2017). Jarak Euclidean didefinisikan pada Persamaan 30.

$$d_i = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (30)$$

Keterangan :

d_i : hasil perhitungan jarak euclidean

p_i : data latih

q_i : data uji

n: jumlah data

3. HASIL DAN PEMBAHASAN

Pada penelitian ini terdapat tiga pengujian utama. Pengujian tersebut meliputi pengaruh nilai k , pengaruh kombinasi fitur dan pengaruh seleksi fitur sehingga nantinya dapat disimpulkan akurasi terbaik dari penelitian ini.

Hasil dari skenario pengujian pertama yaitu pengaruh nilai k , dapat dilihat pada Gambar 11. Hasil pengujian mendapatkan nilai k optimal yaitu $k=1$. Hal ini dikarenakan jika hanya terdapat satu jarak terdekat maka kelas dari data latih langsung diarahkan pada kelas label tanpa terpengaruh ke jarak terdekat yang memiliki kesamaan warna maupun tekstur. Seperti dicontohkan pada Gambar 6 ketika hanya menggunakan fitur warna dimana kedua kelas ini memiliki warna dan nilai fitur yang mirip. Sehingga ketika dilakukan perhitungan jarak, nilai jarak antar kedua kelas ini hampir sama. Jika nilai ketetanggaan lebih dari satu, maka memungkinkan untuk masuk ke kelas miripnya. Dapat dilihat pada Tabel 1. Hasil klasifikasi melabeli data Donat menjadi Genjie Pie, hal ini dikarenakan fitur pada Donat mirip dengan Genjie Pie. Seperti dicontohkan pada fitur deviasi standar H pada data uji. Fitur tersebut dapat dilihat pada Tabel 2, fitur dari data latih deviasi standar H pada kelas Donat dan Genjie Pie pada Tabel 3. Berdasarkan data yang telah

Donat	Genjie pie
-------	------------



Gambar 6. Contoh kelas dengan kemiripan warna

Tabel 1. Hasil klasifikasi dengan $k=3$

Konsisi	Data Uji
$k = 1$	0
$k = 2$	20
$k = 3$	20
Kelas sebenarnya	0
Kelas klasifikasi	20

Tabel 2. Nilai fitur *mean H* data uji

Data uji	Sd_h
1	16.51805

Tabel 3. Nilai fitur *mean H* data latih

Data latih	Donat	Genjie Pie
1	16,95435	16,93331
2	17,19263	17,24879
3	18,25729	17,53506
4	16,64408	17,24879
5	17,37755	17,53506

dianalisis, dapat disimpulkan bahwa nilai fitur deviasi standar H antara citra Donat dan Genjie Pie memiliki nilai yang mirip, sehingga apabila dilakukan ketetanggaan terdekat dengan nilai k lebih dari satu dapat memungkinkan untuk masuk ke kelas yang memiliki warna mirip dengan data uji.

Skenario pengujian yang kedua yaitu pengaruh kombinasi fitur. Hasil pengujian penggunaan fitur HSV dan GLCM dapat dilihat pada Gambar 12. Hasil pengujian mendapatkan hasil klasifikasi yang kurang baik. Hal ini disebabkan kemiripan tekstur antar kelas dan penggunaan fitur GLCM yang terlalu banyak yang menyebabkan setiap fiturnya tidak mencirikan objek dengan dan berpengaruh pada hasil klasifikasi. Disamping itu juga terdapat citra yang memiliki kemiripan warna antar kelas. Contoh kelas yang memiliki kemiripan warna dan tekstur dapat dilihat pada Gambar 7. Penggunaan fitur HSV saja mendapatkan hasil klasifikasi yang terbaik dibandingkan 2 kombinasi fitur lainnya. Hasil pengujian HSV dapat dilihat pada Gambar 13. Dua belas fitur yang digunakan mampu mencirikan objek dengan baik yang dibuktikan dari hasil klasifikasi dan akurasi yang didapatkan. Penggunaan fitur GLCM saja mendapatkan hasil

No	Kelas Awal	Kelas Klasifikasi
1.		
2.		

Gambar 7. Kelas dengan warna dan tekstur mirip

No.	Kelas awal	Kelas klasifikasi
1.		
2.		

Gambar 8. Kelas dengan kemiripan tekstur akurasi terendah dibandingkan dengan dua pengujian sebelumnya.. Contoh citra yang memiliki kemiripan warna dan tekstur sehingga mengalami kesalahan klasifikasi dapat dilihat pada Gambar 11.

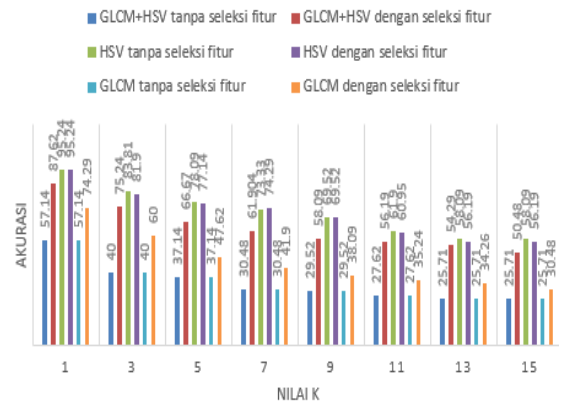
Penggunaan fitur GLCM saja yang menggunakan empat puluh fitur menyebabkan setiap fiturnya tidak mencirikan objek dengan baik karena terdapat banyak kelas yang memiliki kemiripan tekstur serta pengaruh hasil segmentasi yang kurang bagus sehingga membuang informasi penting dari tekstur. Hasil pengujian fitur GLCM dapat dilihat pada Gambar 12. Contoh kelas yang memiliki tekstur mirip dapat dilihat pada Gambar 9 sedangkan contoh kelas yang memiliki hasil segmentasi yang kurang bagus dapat dilihat pada Gambar 10.

No.	Kelas awal	Kelas klasifikasi
1.		
2.		
3.		

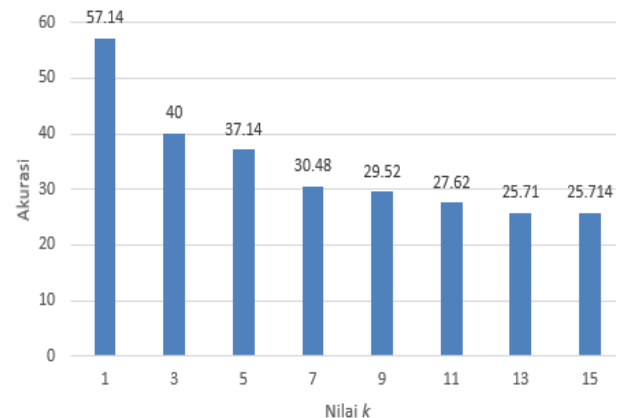
Gambar 9. Kelas dengan kemiripan tekstur

No.	Hasil segmentasi kurang baik
1.	
2.	

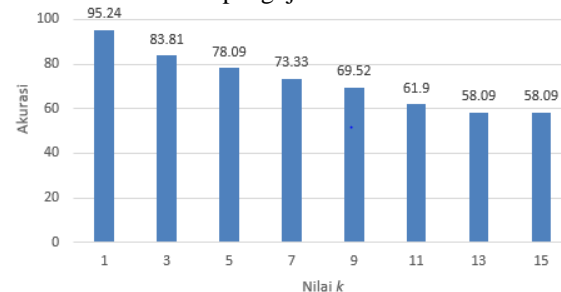
Gambar 10. Hasil segmentasi kurang baik



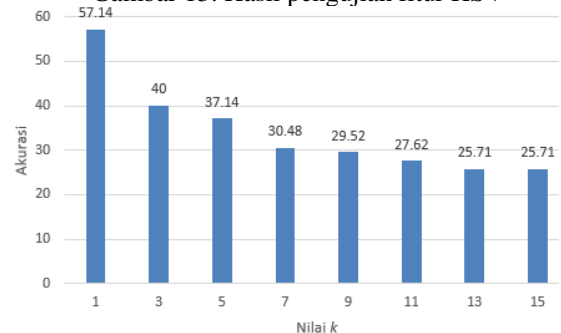
Gambar 11. Hasil pengujian nilai k



Gambar 12. Hasil pengujian fitur HSV dan GLCM



Gambar 13. Hasil pengujian fitur HSV

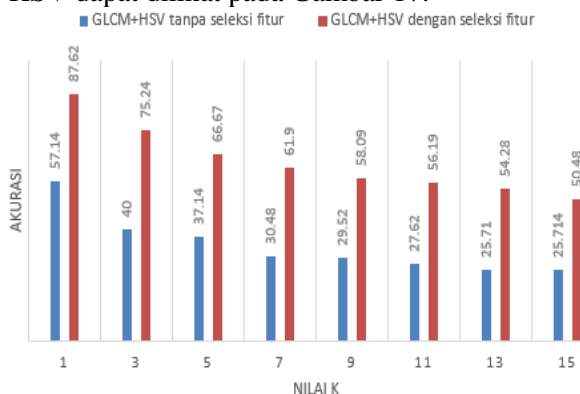


Gambar 14. Hasil pengujian fitur GLCM

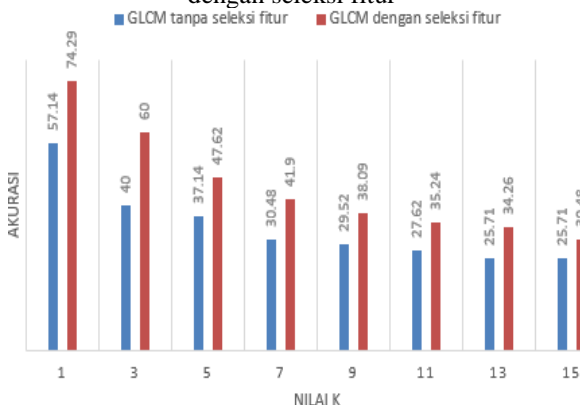
Skenario pengujian yang ketiga, yaitu pengaruh penggunaan *Information Gain*. Penggunaan seleksi fitur *Information Gain* sangat berperan dalam mendapatkan hasil klasifikasi yang baik dan mampu mengurangi beban kerja sistem dengan mengurangi fitur yang banyak namun mampu mendapatkan hasil yang lebih baik dibandingkan tanpa

menggunakan seleksi fitur. Pada pengujian kombinasi fitur HSV dan GLCM dengan mengambil lima belas dari lima puluh dua fitur yang ada, mampu meningkatkan akurasi sebesar 30,48%. Hasil pengujian *Information Gain* menggunakan fitur HSV dan GLCM dapat dilihat pada Gambar 15. Begitu juga ketika menggunakan fitur GLCM saja, seleksi fitur *Information Gain* mampu meningkatkan akurasi sebesar 17,14% dengan mengambil sepuluh fitur dari empat puluh fitur yang digunakan pada penelitian ini.

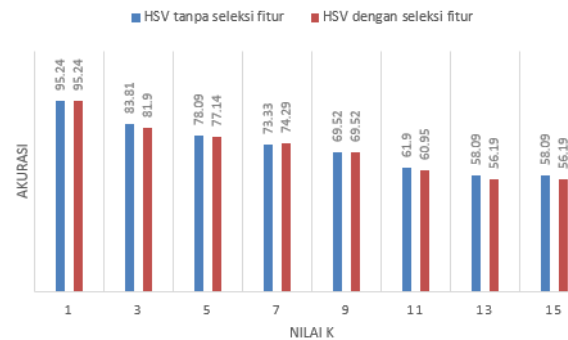
Hasil pengujian *Information Gain* menggunakan fitur GLCM dapat dilihat pada Gambar 15. Sedangkan seleksi fitur *Information Gain* pada *Color Moment* HSV tidak terlalu meningkatkan akurasi dengan signifikan dikarenakan kedua belas fitur *Color Moment* ini berisi informasi penting mengenai informasi objek, sehingga apabila dilakukan pengurangan fitur akan mengurangi informasi objek tersebut. Namun *Information Gain* tetap mampu dalam mengurangi beban kerja sistem dengan mengurangi dua belas menjadi delapan fitur. Hasil pengujian *Information Gain* menggunakan HSV dapat dilihat pada Gambar 17.



Gambar 15. Hasil pengujian HSV dan GLCM dengan seleksi fitur



Gambar 16. Pengujian GLCM dengan seleksi fitur



Gambar 17. Pengujian HSV dengan seleksi fitur

4. PENUTUP

Berdasarkan implementasi, pengujian dan analisis yang telah dilakukan dapat ditarik kesimpulan bahwa penggunaan fitur HSV saja mendapatkan hasil akurasi terbaik sebesar 95,24% dengan nilai $k=1$ baik menggunakan maupun tanpa menggunakan seleksi fitur. *Information Gain* mampu meningkatkan akurasi dari 57,14% menjadi 87,61% pada kombinasi fitur HSV dan GLCM dan 57,14% menjadi 74,28% pada fitur GLCM. *Information Gain* sangat berperan untuk meningkatkan akurasi dan mengurangi beban kerja sistem serta mampu mendapatkan fitur-fitur relevan apabila fitur berjumlah banyak. Namun jika sebaliknya, akan menghasilkan akurasi yang berbeda tipis atau bahkan dapat lebih buruk karena akan membuang informasi penting dari objek yang dicirikan.

Perlu dilakukan pengembangan penelitian dengan menggunakan metode ekstraksi tekstur lain yang lebih mampu mencirikan objek yang memiliki kemiripan dengan objek. Tidak hanya menggunakan fitur warna dan tekstur saja, bisa dengan menambahkan atau mengganti fitur tekstur menjadi fitur bentuk.

5. DAFTAR PUSTAKA

- Alhassan, Khader, A., 2014. Color And Texture Fusion-Based Method For Content Based Image Retrieval. S2. Sudan university of science and technology. Tersedia di <<http://repository.sustech.edu/handle/123456789/10391?show=full>>[diakses 12 November 2018].
- Chormunge, S. & Jena, S., 2016. *Efficient Feature Subset Selection Algorithm for High Dimensional Data*. *International Journal of Electrical and Computer Engineering (IJECE)*, Volume 6, pp. 1880-1888.

- Darwiyanto, E., Binawan, B. P. & Junaedi, D., 2017. Aplikasi GIS Klasifikasi Tingkat Kerawanan Banjir Wilayah Kabupaten Bandung Menggunakan Metode Weighted Product. *Ind. Journal On Computing*, II(1), pp. 59-70.
- Firmansyah & Gatini, T., 2016. Penerapan Metode Content-Bbased Filtering Pada Sistem Rekomendasi Kegiatan Ekstra Kulikuler (Studi Kasus Sekolah ABC). *Jurnal Teknik Informatika dan Sistem Informasi*, Volume 2, pp. 48-56.
- Hastuti, M. T., Widodo, A. W. & Dewi, C., 2018. Identifikasi Kondisi Kesehatan Ayam Petelur Berdasarkan Ciri Warna HSV Dan Gray Level Cooccurrence Matrix (GLCM) Pada Citra Jengger Dengan Klasifikasi K-Nearest Neighbour. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(3), pp. 1054-1062.
- He, Y., Khanna, N., Bushey C. J., & Delp, E. J., 2014. Analysis of Food Images: Features and Classification. *ICIP*, 1(4), pp. 2744-27481.
- Hoonlor, A. et al, 2015. *UCap: A Crowdsourcing Application for the Visually Ipaired and Blind Persons on Android Smartphone*. Chiang Mai, Thailand, International Computer Science and Engineering Conference (ICSEC).
- Jahan, S., H, S. R. & Quadri, S. A., 2018. Bird's Eye Review on Food Image Classification using Supervised Machine Learning. *International Journal of Latest Techology in Engineering, Managements & Applied Science (IJLTEMAS)*, VII(3), pp. 153-159.
- Joutou, T. & Yanai, K., 2009. A Food Image Recognition System With Multiple Kernel Learning. s.l., 16th IEEE International Conference on Image Processing (ICIP).
- Kadir, A., & Susanto. A., 2013. Teori dan Aplikasi Pengolahan Citra. Yogyakarta: penerbit ANDI.
- Kusumanto, R., Pambudi, W. S. & Tompunu, A. N., 2011. Klasifikasi Warna Menggunakan Pengolahan Model Warna HSV. *Jurnal Ilmiah Elite Elektro*, II(2), pp. 83-87.
- Madenda, S., 2015. Pengolahan Citra & Video Digital. Jakarta:Penerbit Erlangga.
- Mohanaiah, P., Sathyanarayana, P. & GuruKumar, L., 2013. Image Texture Feature Extraction Using GLCM Approach. *International Journal of Scientific and Research Publications (IJSRP)*, III(5), pp. 1-5.
- Mwadulo, M. W., 2016. A Review on Feature Selection Methods For Classification Tasks. *International Journal of Computer Applications Technology and Research*, VI(6), pp. 395-402.
- Priambodo, A., Dewi, C. & Triwiratno, A., 2015. Implementasi Metode K-Nearest Neighbour untuk Identifikasi Penyakit Tanaman Jeruk Keprok Berdasarkan Citra Daun. *Doro Journal*, 9(8), p. 56-65.
- Purwati, S., Salimar., Rahayu, S. 2002. Perencanaan Menu Tekanan Darah Tinggi. Penebar Swadaya. Jakarta.
- Puspitasari, R. D., Ratnawati, D. E. & Fauzi, M. A., 2018. Optimasi Susunan Gizi Makanan Bagi Pasien Rawat Jalan Penyakit Jantung Menggunakan Real Coded Genetic Algorithm. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(1), pp. 44-52.
- Redjeki, S., 2013. Perbandingan Algoritma Backpropagation dan K-Nearest Neighbor (K-NN) untuk Identifikasi Penyakit. Jakarta, Seminar Nasional Aplikasi Teknologi Informasi, pp. 1-5.
- Shaltout, N. A., El-Hefnawi, M., Rafea, A. & Moustafa, A., 2014. *Information Gain as a Feature Selection Method for the Efficient Classification of Influenza Based on Viral Hosts*. London, U.K, WCE.
- Singh, S. M. & Hemachandran, K., 2012. Content-Based Image Retrieval using Color Moment Gabor Texture Feature. *International Journal of Computer Science*, IX(5), pp. 299-309.
- Opencv, 2017. Color Conversions. [Online] Doxygen, Tersedia di: https://docs.opencv.org/3.3.0/de/d25/imgproc_color_conversions.htm [Diakses 13 September 2018].