

# SELEKSI FITUR WARNA CITRA DIGITAL BIJI KOPI MENGGUNAKAN METODE *PRINCIPAL COMPONENT ANALYSIS*

*Digital Image Selection of Coffee Seed Using Component Analysis Method*

Danang Aditya Nugraha<sup>1\*</sup>, Anggri Sartika Wiguna<sup>2</sup>

<sup>1,2</sup> Teknik Informatika, Fakultas Sainstek, Universitas Kanjuruhan Malang  
e-mail: \*[danang.adty@unikama.ac.id](mailto:danang.adty@unikama.ac.id), <sup>2</sup> [anggri.sartikawiguna@unikama.ac.id](mailto:anggri.sartikawiguna@unikama.ac.id)

**Abstrak**—Proses klasifikasi citra digital dilakukan terhadap data fitur yang menggambarkan ciri suatu citra. Dalam penelitian sebelumnya mengenai klasifikasi biji kopi untuk menentukan tingkat roasting, digunakan fitur warna RGB dan metode klasifikasi Jaringan Syaraf Tiruan dengan algoritma Back Propagation. Namun dalam penelitian tersebut didapatkan hasil klasifikasi yang kurang memuaskan akibat dari fitur yang digunakan kurang dapat memisahkan ciri dari setiap tingkat roasting biji kopi. Pada penelitian ini dilakukan penambahan fitur warna yang digunakan melalui penggunaan representasi warna RGB dan HSI, yang kemudian dihitung nilai mean, median, standar deviasi, skewness, kurtosis, dan energy pada setiap komponen warna. Dengan demikian terdapat 33 data fitur citra biji kopi. Dengan jumlah fitur yang semakin banyak, menimbulkan masalah baru dimana tidak setiap fitur mempunyai kontribusi yang signifikan dalam proses klasifikasi, sehingga diperlukan proses seleksi fitur untuk mendapatkan fitur dengan kontribusi maksimal. Pada penelitian ini metode Principal Component Analysis diujikan sebagai metode untuk seleksi fitur. Uji coba dilakukan terhadap 240 data citra digital biji kopi, dimana terdapat 120 citra biji kopi Arabika dan 120 citra biji kopi Robusta. Setiap jenis citra biji kopi tersebut terbagi menjadi tingkat roasting mentah, light roasting, medium roasting, dan dark roasting. Hasil penelitian ini didapatkan bahwa rata-rata proses training pada data fitur setelah dilakukan proses seleksi fitur mengalami peningkatan dibandingkan tanpa seleksi fitur. Hal ini terlihat dari 5 kali proses training dengan seleksi fitur didapatkan nilai akurasi terbaik hingga 90,8%, sedangkan tanpa seleksi fitur didapatkan akurasi terbaik adalah 89,6%.

**Kata kunci**— seleksi fitur, jaringan syaraf tiruan, back propagation, principal component analysis

**Abstract**—The image classification process is performed on feature data which representative specific character of an image. In previous studies about classified of coffee beans to determine the level of roasting using RGB color features and artificial neural network classification methods with Back Propagation algorithm, the classified results was unsatisfactory. but In this research, color features was add using RGB and HSI color, and then calculate the average value, median, standard deviation, skewness, kurtosis, and energy on each color component and gave result 33 data features of the image of coffee beans. Added many number of features, can create a new problems, in this case every feature has not a significant contribution in the classification process, so the feature selection process is needed to get maximum result. this study used the Principal Component Analysis method to feature selection. The testing was conducted on 240 digital image data of coffee bean image, which 120 Arabica coffee bean and 120 Robusta coffee bean . Each type of coffee bean image was divided into raw roasting, light roasting, medium roasting, and dark roasting. And the results process of the data features after the selection process was increase, it can be seen from the testing until 5 times process wich gave the best result 90.8%, and without feature selection gave the result 89.6%.

**Keywords**— artificial neural network; Featured Selection; back propagation, principal component analysis

## I. PENDAHULUAN

Pengembangan ilmu dalam bidang teknologi informasi khususnya pada bidang Computer Vision semakin mendapatkan perhatian yang besar, karena pada bidang ini mempelajari bagaimana komputer dapat mengolah data visual untuk mendapatkan suatu informasi yang bermakna. Hal ini semakin diperkuat dengan penerapan computer vision pada berbagai bidang, seperti halnya pada penelitian yang telah dilakukan [1] yang memanfaatkan bidang computer vision untuk melakukan klasifikasi mutu biji kopi, dan juga penelitian [2] yang memanfaatkan teknik dalam bidang computer vision untuk klasifikasi jenis biji kopi Etiopia.

Dalam sebuah sistem computer vision terutama dalam permasalahan klasifikasi, data yang berupa citra digital diolah dalam beberapa tahapan, dalam tahap-tahap proses tersebut tentunya melibatkan sebuah metode kecerdasan buatan sebagai alat dalam pengambilan keputusan. Penerapan metode kecerdasan buatan dalam proses klasifikasi tentunya memerlukan tahap pengolahan data yang lebih spesifik, dikarenakan penggunaan data visual sebagai masukannya. Tahap pengolahan data tersebut bertujuan mendapatkan karakteristik – karakteristik objek yang terkandung dalam citra digital tersebut, yang kemudian karakteristik tersebut digunakan sebagai input dalam proses pengambilan keputusan menggunakan algoritma kecerdasan buatan. Karakteristik tersebut dikenal dengan sebutan fitur objek, dan tahapan tersebut dikenal sebagai tahap ekstraksi fitur.

Beberapa jenis fitur digunakan dalam bidang computer vision, diantaranya adalah fitur bentuk, fitur warna, fitur tekstur. Salah satu fitur paling penting dalam bidang computer vision adalah fitur warna. Dalam sebuah citra digital, terdapat 3 channel warna Red, Green, dan Blue, dari ketiga channel warna tersebut dapat dirubah menjadi bidang warna Hue, Saturation, dan Intensity. Dari kedua bidang warna tersebut dapat dihitung beberapa fitur yang mendasarkan pada perhitungan statistik yaitu nilai mean, median, dan standar deviasi, seperti yang dilakukan pada [3][4]. Berdasarkan penelitian tersebut terlihat bahwa terdapat berbagai jenis fitur warna yang dapat dimanfaatkan untuk proses klasifikasi.

PCA (Principal Component Analysis) merupakan metode transformasi vector data, dan biasanya digunakan untuk mendapatkan gambaran data yang lebih deskriptif dengan dimensi yang lebih sederhana. Beberapa penelitian menggunakan PCA sebagai dasar metode untuk melakukan seleksi fitur, seperti yang dilakukan oleh [5], [6] dan juga [7]. Dalam penelitian sebelumnya [8] telah dilakukan proses klasifikasi untuk menentukan tingkat kematangan roasting biji kopi, dan pada permasalahan ini hanya melibatkan fitur warna RGB. Berdasarkan hasil penelitian tersebut terlihat proses klasifikasi kurang optimal dikarenakan fitur data yang terlalu sedikit. Sehingga pada penelitian ini dilakukan penambahan fitur warna dengan melibatkan warna RGB dan HSI yang kemudian dihitung nilai *mean*, *median*, standar deviasi, *skewness*, *kurtosis*, dan *energy*. Fitur yang semakin banyak tentunya harus melalui proses seleksi untuk memilih fitur dengan kontribusi optimal dalam proses klasifikasi. Oleh karena itu pada penelitian ini akan dilakukan pengujian metode Principal Component Analysis untuk melakukan seleksi fitur dalam permasalahan klasifikasi tersebut.

## II. LANDASAN TEORI

### 2.1 Ekstraksi Fitur

suatu perhitungan untuk memperoleh suatu nilai yang mewakili karakteristik tertentu pada suatu citra digital. Fitur yang banyak dimanfaatkan pada bidang *computer vision* diantaranya adalah fitur bentuk, warna dan tekstur. Fitur warna merupakan fitur yang penting dikarenakan Penyelesaian permasalahan dalam bidang *computer vision* melibatkan beberapa tahap proses. Tahap proses yang penting diantaranya adalah tahap ekstraksi fitur. Pada tahap ini dilakukan penggunaannya yang sangat intensif dalam bidang *computer vision*.

#### 2.1.1 Fitur Warna

Dalam sebuah citra digital, warna yang ditampilkan merupakan perwakilan nilai intensitas cahaya yang ditangkap oleh sensor pada proses akuisisi citra digital. Pada citra digital berwarna secara mendasar terdapat gabungan tiga nilai intensitas warna yaitu *Red*,

Green dan Blue. Fitur warna merupakan bentuk hasil pengolahan terhadap nilai-nilai intensitas ini yang nantinya dapat mewakili suatu karakteristik warna pada citra digital tersebut[9][10].

Perhitungan statistik merupakan cara yang umum digunakan untuk mengolah nilai intensitas piksel, dalam rangka mendapatkan suatu karakteristik tertentu dari sebuah citra digital. Pada penelitian yang telah dilakukan [11] digunakan beberapa perhitungan statistik sebagai metode untuk mendapatkan karakteristik warna tertentu dari sebuah citra digital beberapa perhitungan tersebut adalah :

- *Mean*

Perhitungan ini menghitung rata-rata intensitas pixel pada setiap channel warna, perhitungan ini bertujuan untuk melihat tingkat kecerahan citra pada channel warna tersebut. Perhitungan mean ditunjukkan pada persamaan berikut

$$mean = \sum_{i=1}^L r_i P(r_i) \quad (1)$$

- *Median*

Pada perhitungan ini bertujuan untuk mengambil nilai tengah intensitas pixel pada setiap channel warna

- *Standar Deviasi*

Perhitungan standar deviasi menjadi fitur yang menggambarkan tingkat kontras nilai intensitas pada setiap channel warna. Perhitungan standar deviasi ditunjukkan pada persamaan berikut

$$Stddev = \sqrt{\sum_{i=1}^L (r_i - mean)^2 P(r_i)} \quad (2)$$

- *Skew*

Perhitungan *Skew* bertujuan untuk melihat kecenderungan histogram intensitas warna, perhitungan ini mengukur nilai ketidakberagaman pada distribusi intensitas piksel disekitar nilai *Mean*. Perhitungan *skew* ditunjukkan oleh persamaan berikut ini

$$Skew = \frac{1}{(stddev)^3} \sum_{i=1}^L (r_i - mean)^3 P(r_i) \quad (3)$$

- *Kurtosis*

Perhitungan ini berfungsi untuk mengukur keadaan puncak distribusi intensitas pada nilai *mean*. Apabila bernilai tinggi menunjukkan puncak distribusi yang tajam, sedangkan nilai *kurtosis* rendah menunjukkan puncak distribusi

yang tumpul. Perhitungan *kurtosis* ditunjukkan oleh persamaan sebagai berikut

$$Kurtosis = \frac{1}{(stddev)^4} \sum_{i=1}^L (r_i - mean)^4 P(r_i) \quad (4)$$

- *Energy*

Perhitungan ini berfungsi untuk mengukur *uniformity* distribusi nilai intensitas piksel. Nilai *energy* yang tinggi menunjukkan bahwa nilai piksel terdistribusi pada sebagian kecil intensitas saja

$$energy = \sum_{i=1}^L P(r_i)^2 \quad (5)$$

$$R(x, y) = \begin{cases} 1 & \text{jika } f(x, y) \geq T \\ 0 & \text{sebaliknya} \end{cases} \quad (5)$$

## 2.2 Principal Component Analysis(PCA)

*Principal Component Analysis* adalah suatu metode untuk mengidentifikasi pola dalam suatu data dan juga untuk menonjolkan adanya perbedaan ataupun kesamaan di dalam suatu kumpulan data. Metode ini biasa digunakan sebagai alat untuk mereduksi dimensi data, menjadi bentuk yang berada pada bidang nilai yang berbeda. Metode ini bekerja dengan cara menghitung *covariance matrix* dari data, dan kemudian mencari *eigenvectors*, dan *eigenvalues*.)

Penelitian yang dilakukan oleh [5] memaparkan sebuah cara memanfaatkan metode PCA sebagai cara untuk melakukan seleksi terhadap fitur yang digunakan untuk melakukan pengenalan wajah. Proses seleksi fitur dengan menggunakan metode PCA sesuai dengan urutan langkah sebagai berikut

1. Menghitung *covariance matrix* dari atribut data fitur yang telah dikurangi dengan nilai *mean* nya
2. Berdasarkan *covariance matrix* yang telah dihasilkan, kemudian dihitung *eigenvectors* dan *eigenvalues*. Dari daftar *eigenvectors* pilih sejumlah *m* nilai *eigenvalues* terbesar, kemudian simpan *eigenvectors* terpilih sebagai  $V_1, \dots, V_m$
3. Menghitung kontribusi setiap fitur dengan mengkalulasi nilai setiap fitur pada *eigenvectors* yang sudah dipilih sesuai dengan persamaan 6.

$$c_j = \sum_{p=1}^m |V_{pj}| \quad (6)$$

4. Memilih sejumlah nilai  $C_j$  terbesar sesuai dengan jumlah fitur yang ingin dipertahankan, sehingga didapatkan fitur ke- $j$  merupakan fitur yang signifikan..

### III. METODE

Proses penelitian dilakukan dalam beberapa tahap. Tahap pertama adalah pengembangan aplikasi untuk melakukan pengolahan awal terhadap citra data kopi Aplikasi yang dikembangkan memiliki fungsionalitas sebagai berikut:

- a. Proses pengubahan citra digital berwarna menjadi citra *grayscale* yang kemudian dilakukan proses segmentasi
- b. Melakukan beberapa proses ekstraksi fitur warna RGB dan HSI pada setiap citra uji dan melakukan penyimpanan fitur tersebut, untuk dilakukan seleksi dan kemudian melalui proses klasifikasi sesuai dengan target kelas yang telah ditentukan
- c. Proses seleksi fitur menggunakan metode PCA untuk mendapatkan fitur yang signifikan

Hasil dari aplikasi diatas berupa data fitur awal yang tidak melalui seleksi terdiri dari 33 fitur citra dan data hasil proses seleksi fitur dimana dilakukan seleksi untuk mendapatkan 20 fitur. Kedua data ini kemudian digunakan dalam proses klasifikasi menggunakan metode Jaringan Syaraf Tiruan dengan algoritma pembelajaran Back Propagation. Dikarenakan dalam proses pembelajaran metode ini dapat menghasilkan hasil *training* yang berbeda tergantung inisialisasi pada awal proses, maka dilakukan 5 kali proses training, testing dan validasi pada kedua data fitur sebagai perbandingan.

Data citra digital yang digunakan dalam penelitian ini terdiri dari 240 data citra yang terbagi menjadi 120 citra kopi Arabika dan 120 citra kopi robusta. Setiap citra telah dikelompokkan dalam 4 kelas tingkat *roasting* yaitu Mentah, *Light Roasting*, *Medium Roasting*, dan *Dark Roasting*. Dari 240 data tersebut yang telah diubah menjadi fitur citra kemudian dibagi menjadi *data training*, *data validation*, dan *data testing* dengan prosentase 70%:15%:15%.

### IV. HASIL

Tahap pertama dalam proses pengujian adalah menguji proses pengolahan citra dan tahap

ekstraksi fitur. Data berupa citra digital RGB diubah dalam warna HSI, yang kemudian juga diubah menjadi citra grayscale untuk proses segmentasi. Berikut contoh hasil segmentasi citra untuk mendapatkan area biji kopi pada citra terlihat pada Gambar 1.

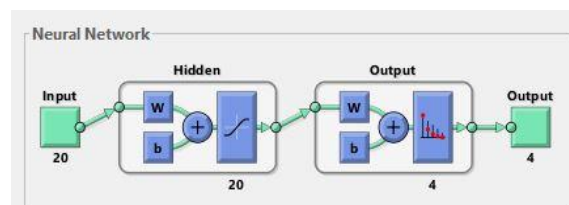


Gambar 1. Citra input dan hasil segmentasi

Citra yang telah melalui proses segmentasi kemudian diambil bagian area biji kopi untuk didapatkan representasi warna RGB dan HSI. Nilai warna tersebut kemudian dihitung nilai *mean*, *median*, standar deviasi, *skewness*, *kurtosis*, dan *energy*. Sehingga total terdapat 33 fitur yang mewakili setiap data citra. Proses ini dilakukan pada setiap data citra kemudian disimpan pada sebuah file.

File tersebut diduplikasi, agar nantinya dapat dibandingkan hasil klasifikasi pada 33 fitur data, dan hasil klasifikasi pada data yang telah melalui seleksi. Proses seleksi dilakukan dengan menggunakan metode PCA, dan ditentukan proses seleksi dilakukan untuk mendapatkan 20 fitur dengan kontribusi paling signifikan.

Kedua data fitur tersebut kemudian digunakan untuk proses klasifikasi dengan menggunakan Jaringan Syaraf Tiruan dengan konfigurasi jaringan seperti pada Gambar 2. Untuk komposisi data training, data testing, dan data validasi, yaitu 70%, 15% dan 15 %.



Gambar 2. Konfigurasi Jaringan Syaraf Tiruan

Karena hasil proses klasifikasi dengan menggunakan Jaringan Syaraf Tiruan dapat berubah-ubah bergantung dari inisialisasi

proses, maka dilakukan 5 kali pengujian klasifikasi pada kedua data fitur dan hasil perbandingan akurasi dapat dilihat pada Tabel 1 berikut ini.

Tabel 1. Perbandingan Akurasi Klasifikasi

No	Fitur Awal			Fitur Hasil seleksi		
	Trainin g	Testin g	Valida si	Trainin g	Testin g	Valida si
1	92,3%	77,8%	88,9%	90,5%	80,6%	94,4%
2	89,9%	86,1%	86,1%	90,5%	97,2%	86,1%
3	86,9%	91,7%	94,4%	90,5%	94,4%	88,9%
4	89,3%	88,9%	88,9%	91,1%	86,1%	94,4%
5	88,7%	91,7	88,9%	89,3%	97,2%	91,7%

Hasil pada Tabel 1 menunjukkan bahwa hasil klasifikasi pada fitur yang telah mengalami seleksi memiliki akurasi hasil klasifikasi lebih baik terutama pada hasil tahap testing. Untuk total akurasi terbaik secara keseluruhan dapat dilihat hasil *confusion matrix* pada Gambar 3 dan Gambar 4.



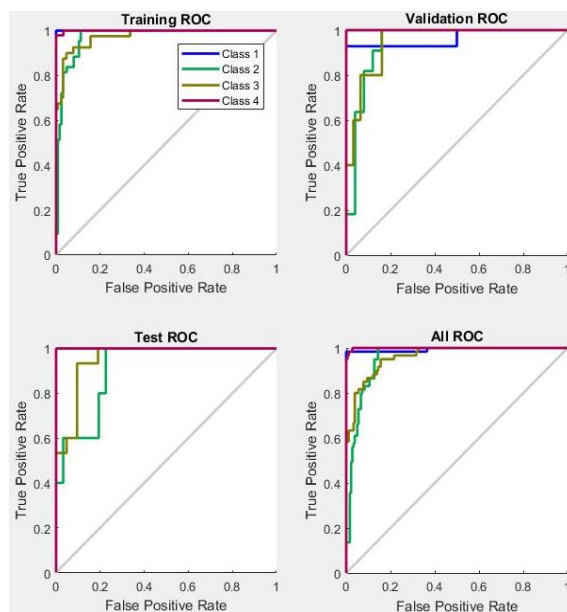
Gambar 3. *Confusion Matrix* hasil klasifikasi tanpa seleksi



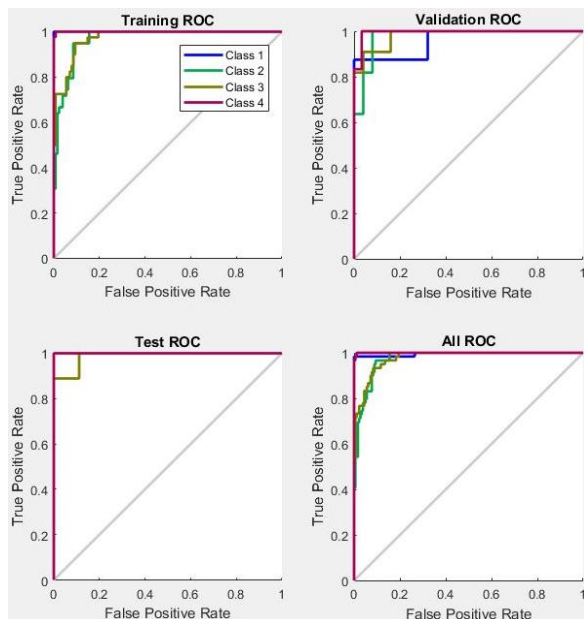
Gambar 4. *Confusion Matrix* hasil klasifikasi terbaik dengan seleksi PCA

Berdasarkan perbandingan diatas terlihat bahwa nilai akurasi dari proses klasifikasi pada data yang telah mengalami seleksi lebih tinggi dibandingkan apabila data fitur tidak melalui seleksi. Sehingga hal ini menunjukkan tidak semua fitur memiliki kontribusi yang signifikan bagi proses klasifikasi, bahkan ada yang berkontribusi negatif.

Sedangkan hasil kurva ROC dari klasifikasi terbaik dari dua macam data fitur dapat dilihat pada Gambar 5 dan Gambar 6.



Gambar 5. Kurva ROC tanpa seleksi



Gambar 6. Kurva ROC dengan seleksi PCA

Berdasarkan kurva ROC menunjukkan bahwa permasalahan terberat dalam proses klasifikasi ini adalah membedakan citra dari biji kopi *light roasting* dan *medium roasting*. Terlihat dari hasil kurva kelas 2 dan 3 yang selalu menjauhi area pojok sebelah kiri atas. Akan tetapi dari kurva ROC ini juga terlihat bahwa perbandingan antara klasifikasi dengan seleksi fitur memiliki performa yang lebih baik dibandingkan klasifikasi pada data fitur tanpa ada seleksi.

#### V. KESIMPULAN

Berdasarkan hasil ujicoba yang telah dilakukan terhadap dua macam kumpulan fitur citra digital dapat diambil kesimpulan bahwa penambahan fitur jelas memberikan hasil klasifikasi yang lebih baik. Tetapi tidak semua fitur yang ditambahkan memberikan kontribusi yang signifikan pada proses klasifikasi. Hal ini terlihat dari hasil akurasi proses klasifikasi yang lebih baik apabila dilakukan pada data fitur yang telah melalui proses seleksi dengan menggunakan metode PCA

#### DAFTAR PUSTAKA

- [1] G. O. F. Parikesit dan Ferdiansjah, "Coffee Bean Grade Determination Based on Image Parameter," *TELKOMNIKA*, vol. 9, no. 3, hal. 547–554, Des 2011.
- [2] B. Turi, G. Abebe, dan G. Goro, "Classification of Ethiopian Coffee Beans Using Imaging Technique," *East African J. Sci.*, vol. 7, no. 1, hal. 1–10, 2013.
- [3] R. Venkata Ramana Chary, "Feature Extraction Methods For Color Image Similarity," *Adv. Comput. An Int. J.*, vol. 3, no. 2, hal. 147–157, Mar 2012.
- [4] M. K. Slamet Widodo, "QUALITY EVALUATION OF MELINJO SEEDS (*Gnetum gnemon L.*) USING DIGITAL IMAGE PROCESSING," *J. Tek. Pertan. Lampung*, vol. 7, no. 2, hal. 106–14, 2018.
- [5] F. Song, Z. Guo, dan D. Mei, "Feature Selection Using Principal Component Analysis," in *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, 2010, hal. 27–30.
- [6] Y. Lu, I. Cohen, X. Sean Zhou, dan Q. Tian, "Feature Selection Using Principal Feature Analysis," *ACM Multimedia*, Sep 2007.
- [7] M. Morchid, R. Dufour, P.-M. Bousquet, G. Linares, dan J.-M. Torres-Moreno, "Feature selection using Principal Component Analysis for massive retweet detection," *Pattern Recognit. Lett.*, vol. 49, hal. 33–39, Nov 2014.
- [8] D. A. Nugraha dan A. S. Wiguna, "Klasifikasi Tingkat Roasting Biji Kopi Menggunakan Jaringan Syaraf Tiruan Backpropagation Berbasis Citra Digital," *SMARTICS J.*, vol. 4, no. 1, hal. 1–4, Sep 2018.
- [9] M. E. Purbaya, "Penerapan Model Generatif Pada Kerangka Kerja Deep Learning Untuk Menerjemahkan Citra Sketsa Daun Menjadi Citra Alami Daun," *Res. Comput. Inf. Syst. Technol. Manag.*, vol. 1, no. 2, hal. 39–45, 2018.
- [10] M. H. Abdurrahman, E. Suhartono, dan E. Wulandari, "Deteksi Kualitas Kemurnian Susu Sapi Melalui Pengolahan Citra Digital Menggunakan Metode Scale Invariant Feature Transform ( Sift ) Dan Klasifikasi K-Nearest Neighbor ( Knn ) Quality Detection of Cow ' S Milk Purity Using Digital Image Processing Method O," in *e-Proceeding of Engineering*, 2019, vol. 6, no. 2, hal. 3845–3852.

- [11] P. Kothyari dan S. Dwivedi, "Content Based Image Retrieval using Statistical Feature and Shape Extraction," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 4, no. 6, Jun 2016.