

## Klasifikasi Jenis Buah Mangga Menggunakan Metode *K-Means Clustering*

Agyztia Premana<sup>1\*</sup>, Akhmad Pandhu Wijaya<sup>2</sup>

<sup>1</sup>Teknik Informatika, Universitas Muhadi Setiabudi, Brebes 52252, Indonesia

<sup>2</sup>Teknik Informatika, Universitas Wahid Hasyim, Semarang 50236, Indonesia

email: a.premana@umus.ac.id<sup>1</sup>, pandhuds@unwahas.ac.id<sup>2</sup>

### Abstrak

Perkembangan teknologi saat ini semakin pesat yang dampaknya sangat memudahkan masyarakat dalam banyak bidang. Salah satu bidang yang terkena dampaknya ialah perindustrian. Di Indonesia sendiri mangga merupakan salah satu jenis buah yang paling banyak diproduksi. Mangga dapat diolah menjadi berbagai jenis makanan, seperti selai, keripik, roti dan lain sebagainya. Berbeda hasil olahan berbeda pula jenis mangga yang digunakan. Salah satu contoh efek perkembangan teknologi saat ini yang berdampak pada bidang perindustrian ialah memungkinkan adanya sistem yang mampu belajar sendiri seperti manusia metode ini dinamakan dengan Jaringan Syaraf Tiruan. Sistem yang akan dibangun ini menerapkan jaringan syaraf tiruan untuk pemodelannya dan menggunakan ekstraksi ciri berupa mean rgb dan standar deviasi rgb, perimeter, luas, panjang, lebar, kebulatan, dan kerampingan. Berdasarkan hasil percobaan, Waktu Komputasi yang dibutuhkan untuk proses ekstraksi fitur untuk setiap dataset adalah rata-rata sebesar 0,85 detik, sedangkan waktu komputasi untuk uji coba pada data uji adalah rata-rata sebesar 0,006 detik. Dapat disimpulkan bahwa waktu komputasi untuk setiap dataset adalah rata-rata sebesar 0,856 detik.

**Kata kunci:** Klasifikasi, Mangga, Image Processing, Backpropagation.

### Abstract

*Today's technological advancement is quick and simple for those working in a variety of sectors. Industry inventions have an impact on one of these sectors. Mangoes are one of the fruit varieties that are most extensively cultivated in Indonesia. Mangoes can be turned into many different kinds of food, including jam, chips, bread, and others. The variety of mango used also affects how the products are processed. The potential for a system that can self-learn via artificial neural networks, similar to the early human technique, is one example of contemporary technical advancements that have an impact on the industrial sector. The system to be constructed uses an artificial network for modeling and rgb mean and rgb standard feature extraction for feature extraction. in the form of rgb mean and rgb standard deviation, perimeter, area, length, width, roundness, and slenderness. Based on the experiment, the computational time required for the feature extraction process for each dataset is an average of 0.85 seconds, while the computational time required for testing on the test data is an average of 0.006 seconds. It can be said that the computation time for each dataset is an average of 0.856 seconds.*

**Keywords:** Classification, Mango, Image Processing, Backpropagation

## 1. PENDAHULUAN

Perkembangan teknologi membuat manusia ingin meningkatkan efektifitas dan efisiensi dalam berbagai bidang, salah satunya dalam bidang pertanian. Dalam dunia pertanian, kemajuan teknologi sangat dibutuhkan untuk menunjang kegiatan-kegiatan yang ada pada bidang pertanian, salah satunya dalam pengolahan hasil pertanian dan perkebunan. Industri pengolahan hasil pertanian dan perkebunan kini semakin berkembang pesat seiring perkembangan teknologi yang ada

khususnya untuk produksi Buah Mangga (Lee et al. 2017). Di Negara Indonesia sendiri mangga banyak disukai oleh semua kalangan untuk di konsumsi secara langsung sebagai buah atau diolah menjadi produk konsumsi lain seperti, roti mangga, selai mangga dan lain sebagainya. Hal ini dapat dilihat dari produksi mangga di Indonesia terus meningkat. Menurut Badan Pusat Statistik dan Direktorat Jenderal Hortikultura, produksi mangga 3 tahun terakhir ini terus mengalami perkembangan. Meskipun banyaknya produksi mangga di Indonesia,

namun tidak seimbang dengan pengolahan terhadap Buah mangga tersebut. Penelitian ini akan mencoba membantu pihak pengolahan mangga yang masih sedikit, namun powerful dalam pengolahan manga (Ristian 2017). Pada proses pengolahan Buah mangga tersebut, ada beberapa tahapan sebelum produk yang dihasilkan di distribusikan ke konsumen. Salah satu tahapannya adalah pemilihan mangga hasil pertanian dan perkebunan sesuai dengan kebutuhan untuk diolah. Pemilihan jenis Buah mangga sangat perlu dilakukan, sebab berbeda jenis mangga berbeda pula cara pengolahan dan hasil olahannya.

Proses pemilihan Buah mangga dilakukan berdasarkan bentuk, ukuran, dan warna pada Buah Mangga. Pemilihan warna meliputi mean rgb dan standar deviasi rgb, ukuran meliputi perimeter, luas, dan lebar, sedangkan bentuk meliputi kebulatan dan kerampingan. Tujuan dari proses pemilihan Buah mangga ini untuk memisahkan Buah mangga berdasarkan jenisnya. Petani atau penyeleksi Buah mangga umumnya mengidentifikasi jenis mangga dapat dilihat dari ukuran dan warna kulit nya saja, karena hal tersebut yang paling mudah untuk dilakukan (Malang 2016).

Walaupun mudah dilakukan, pada kenyataannya hal tersebut memiliki permasalahan yang terkadang membuat proses pemilihan jenis Buah mangga kurang optimal, terlebih jika hal tersebut dilakukan secara manual. Hal tersebut bisa terjadi karena perbedaan persepsi dari masyarakat tentang penyeleksi Buah mangga terhadap faktor komposisi warna, bentuk, ukuran citra suatu objek berbeda-beda walaupun objek yang dilihat sama persis. Selain itu, dengan semakin besarnya produksi mangga di Indonesia, maka seharusnya semakin banyak pula perusahaan atau pabrik untuk mengolah mangga tersebut. Sebuah pabrik ataupun perusahaan pasti akan mengolah mangga dalam skala yang sangat besar. Jika pabrik atau perusahaan tersebut menggaji karyawan untuk memilih jenis mangga, maka pabrik tersebut harus menggaji banyak pegawai hanya untuk memperhatikan jenis mangga secara manual (Utami et al. 2017).

Berdasarkan permasalahan tersebut diatas, maka perlu diadakan penelitian tentang implementasi aplikasi untuk mengidentifikasi jenis Buah mangga dengan memanfaatkan pengolahan citra

digital (image processing). Penelitian ini dilakukan karena dapat menjadi salah satu alternatif teknologi untuk dapat membantu mengurangi atau mengatasi permasalahan yang ada. Membangun sebuah aplikasi untuk mengidentifikasi jenis Buah mangga perlu mengkombinasikan beberapa bidang ilmu seperti pengolahan citra digital (image processing) dan Jaringan Saraf Tiruan (K-MEANS). Pengolahan citra digital (image processing) merupakan teknik untuk mengolah citra suatu objek secara digital yang mana setelah melalui proses pengolahan akan didapatkan suatu informasi yang dapat diolah oleh komputer untuk berbagai kebutuhan (Sari, Hidayat, and Atmaja 2016)

Hal ini nantinya dapat diterapkan pada saat mengolah citra digital dari bentuk, ukuran, dan warna mangga dalam klasifikasi jenis Buah Mangga. Untuk mengidentifikasi objek citra digital Buah mangga dengan pengelompokan jenis Buah mangga perlu bantuan bidang ilmu yang lainnya yaitu Jaringan Saraf Tiruan (K-MEANS). Secara umum K-MEANS merupakan suatu bidang ilmu yang mempelajari tentang teknik mengolah data agar dapat digunakan untuk memberikan suatu indikasi tertentu yang bermanfaat. Dalam K-MEANS terdapat banyak teknik atau metode dalam mengolah suatu data salah satunya Backpropagation (Kim and So 2018).

Penelitian ini nantinya diharapkan dapat memberikan solusi terhadap permasalahan yang ada pada pengolahan hasil pertanian dan perkebunan. Khususnya pada proses pemilihan Buah mangga berdasarkan tingkat kematangannya. Sehingga pengimplementasian aplikasi pada penelitian ini dapat membantu dalam pengolahan hasil pertanian dan perkebunan sesuai kebutuhan konsumen secara cepat, tepat dan efisien.

## 2. METODE

Pada bab ini diuraikan mengenai rencana langkah-langkah yang akan dilakukan dalam penelitian. Termasuk bagaimana cara mendapatkan data, cara mengolah data, cara membuat alat uji jika diperlukan, cara analisis data, cara pengujian, dan di bagian akhir diberikan desain alat uji serta *user interface* dari alat uji yang akan dibangun. Clustering data dapat dibedakan menjadi dua tujuan, yaitu clustering untuk pemahaman dan clustering untuk penggunaan. Jika tujuan untuk

pemahaman maka cluster yang terbentuk harus menangkap struktur alami data (Zhang et al. 2021).

Biasanya proses clustering dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan ini seperti summarization (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi, dan sebagainya. Sementara jika tujuannya untuk penggunaan, biasanya tujuan utama untuk mencari prototype cluster yang paling representative terhadap data dan memberikan abstraksi dan setiap objek data dalam cluster di mana sebuah data terletak didalamnya (Younesi and Amirani 2017).

**2.1. Buah Mangga**

Pengembangan dari sisi tampilan dapat ditambah untuk memudahkan membaca grafik sehingga pengguna dapat melihat bagian grafik dari akurasi dan waktu yang dihasilkan. Mangga (*Nephelium lappaceum*) adalah buah yang berasal dari Asia Tenggara, seperti Indonesia. Bila diperhatikan, mangga tampak mirip dengan buah leci. Meski sama-sama berwarna merah dan terasa segar, buah mangga memiliki rambut yang tumbuh pada kulit buahnya, sedangkan buah leci tidak. Selain itu, kandungan gizi dan vitamin dalam buah mangga pun berbeda.

Tabel 1. Kandungan Gizi dan Vitamin

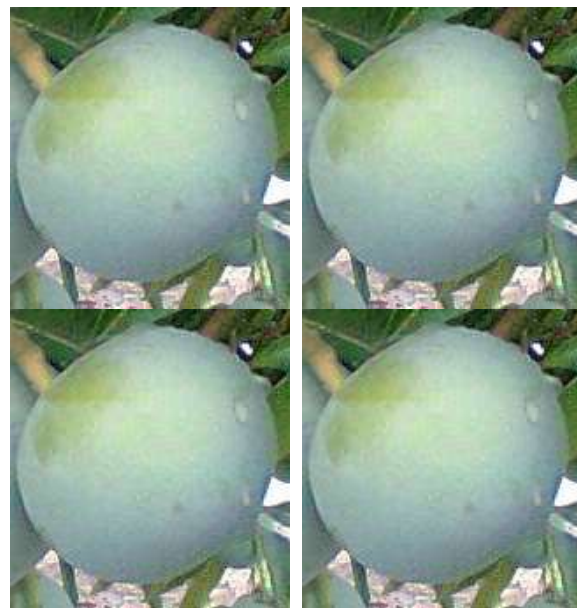
Nama	Kandungan
Protein	0,3g
air	86,6g
Lemak	0,1g
Karbohidrat	12,3g
Serat	1,6g
Kalsium	20mg
Fosfor	16mg
Zat Besi	0,1g

Kandungan nutrisi yang kaya pada buah mangga ini menawarkan segudang khasiat yang baik bagi kesehatan. Mangga termasuk jenis buah yang baik untuk sistem pencernaan. Pasalnya, sekitar setengah daging buah mangga mengandung serat larut tidak air. Artinya, jenis serat ini tidak menyatu dengan air dan langsung melewati sistem pencernaan. Sebagian besar serat tidak larut ini langsung masuk ke sistem pencernaan dan mendorong feses di dalam usus. Hal ini yang akan membantu melancarkan pencernaan dan tidak sulit mengeluarkan feses saat BAB. Sementara itu, kandungan jenis serat larut air

pada buah mangga akan menjadi makanan bagi bakteri usus. Dengan begitu, usus menghasilkan asam lemak rantai pendek, seperti asetat, propionat, dan butirrat, sebagai makanan untuk sel usus (Parida and Bhoi 2017).

**2.2. Citra Biner**

Citra biner (binary image) adalah citra yang hanya mempunyai dua nilai derajat keabuan yaitu hitam dan putih. Meskipun saat ini citra berwarna lebih disukai, namun pada beberapa aplikasi citra biner masih dibutuhkan, misalnya pada pengolahan citra dalam menentukan jenis mangga seperti yang akan dibahas dalam penelitian ini. Citra biner hanya mempunyai dua nilai derajat keabuan yaitu hitam dan putih. Pixel-pixel objek bernilai 1 dan pixel-pixel latar belakang bernilai 0. Pada saat menampilkan gambar, 0 adalah putih dan 1 adalah hitam. Jadi pada citra biner, latar belakang berwarna putih sedangkan objek berwarna hitam. Sebagai contoh gambar dibawah ini dengan threshold 0,7 (Nnolim 2018).



Gambar 1. Citra Buah Mangga.

**2.3. Citra RGB**

Citra berwarna atau biasa dinamakan citra RGB, merupakan jenis citra yang menyajikan warna dalam bentuk komponen R (merah), G (green), B (biru). Setiap komponen warna menggunakan delapan bit (nilainya berkisar antara 0 sampai dengan 255). Dengan demikian kemungkinan warna yang dapat disajikan mencapai 255 x 255 x 255 atau 16.581.375 warna (Abdul dan Adhi, 2013). Table 2.4 menunjukkan contoh warna dan

nilai R, G, B (Sumaiya and Shantha Selva Kumari 2017).

Tabel 3. Warna dan Nilai

Warna	R	G	B
Hijau	255	0	0
Merah	0	255	0
Biru	0	0	255
Hitam	0	0	0
Putih	255	255	255

Model RGB biasa disajikan dalam bentuk kubus tiga dimensi, dengan warna merah, hijau, warna hitam berada pada titik asal dan warna putih berada di sumbu kubur yang bersebrangan. Fitur warna dapat diperoleh melalui perhitungan statistic seperti rerata dan deviasi standar. Sebagai contoh fitur-fitur tersebut dapat digunakan untuk kepentingan identifikasi maupun klasifikasi. Perhitungan dikenakan pada setiap komponen R, G, dan B (Shokrzade et al. 2021).

**2.4. K-Means Klustering**

Metode clustering menurut strukturnya dibagi menjadi dua yaitu pengelompokan hirarki dan partitioning. Pengelompokan hirarki memiliki aturan satu data tunggal bisa dianggap sebagai sebuah kelompok, dua atau lebih kelompok kecil dapat bergabung menjadi satu kelompok besar dan begitu seterusnya hingga semua data dapat bergabung menjadi satu kelompok.

Metode clustering hirarki merupakan satu-satunya metode yang masuk ke dalam kategori pengelompokan hirarki. Metode clustering partitioning membagi set data ke dalam sejumlah kelompok yang tidak tumpang tindih (overlap) antara satu kelompok dengan kelompok yang lain artinya setiap data hanya menjadi anggota satu kelompok.

Metode seperti K-Means dan DBSCAN masuk dalam kategori pengelompokan partitioning. Metode clustering menurut strukturnya dibagi menjadi dua yaitu pengelompokan hirarki dan partitioning. Pengelompokan hirarki memiliki aturan satu data tunggal bisa dianggap sebagai sebuah kelompok, dua atau lebih kelompok kecil dapat bergabung menjadi satu kelompok besar dan begitu seterusnya hingga semua data dapat bergabung menjadi satu kelompok. Metode clustering hirarki merupakan satu-satunya metode yang masuk ke dalam kategori pengelompokan hirarki (Alamuru and Jain 2021).

$$d_{ik} = \sqrt{\sum_{j=1}^m (X_{ij} - C_{kj})^2}$$

Metode clustering partitioning membagi set data ke dalam sejumlah kelompok yang tidak tumpang tindih (overlap) antara satu kelompok dengan kelompok yang lain artinya setiap data hanya menjadi anggota satu kelompok. Metode seperti K-Means dan DBSCAN masuk dalam kategori pengelompokan partitioning.

**2.5. Fitur Ekstraksi**

Dari semua penjelasan diatas dapat, fitur-fitur diatas dapat diturunkan menjadi beberapa fitur lagi, diantaranya.

**Kebulatan Bentuk**

Kebulatan bentuk adalah perbandingan antara luas objek dan kuadrat Perimeter yang dinyatakan dengan rumus seperti berikut.

$$Kebulatan = 4\pi \frac{A(R)}{P^2(R)}$$

Hasilnya berupa nilai  $\leq 1$ . Nilai 1 menyatakan objek R berbentuk lingkaran. Terkadang, future ini dinamakan kekompakan.

**Kerampingan Bentuk**

Kerampingan bentuk adalah perbandingan antara lebar dan panjang yang dinyatakan dengan rumus sebagai berikut.

$$Kerampingan = \frac{lebar}{panjang}$$

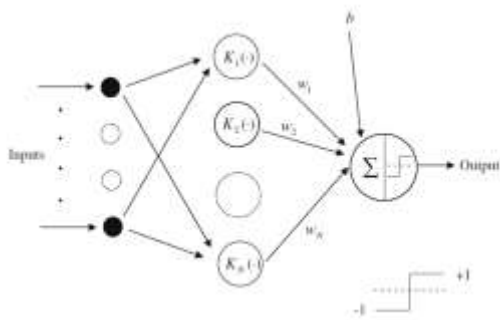
Panjang adalah panjang objek dan lebar adalah lebar objek. Fitur ini terkadang disebut rasio aspek.



Gambar 7. Gambaran Klasifikasi

### 3. HASIL DAN PEMBAHASAN

Pada penelitian ini dipilih gambar yang mempunyai kerumitan tinggi untuk menguji performa sekaligus akurasi metode *K – Means*. Metode *K – Means* menempuh beberapa tahap dalam proses klasifikasi teks, yaitu tahap membuat struktur transformasi warna secara spesifik, dan tahap klasifikasi, pada tahap pelatihan dilakukan proses analisis terhadap sampel gambar buah mangga.



Gambar 5. Hasil Clusterin Pixel Buah Mangga

Pada proses klasifikasi pengujian menggunakan hasil ekstraksi ciri data pengujian dan hasil proses klasifikasi pelatihan. Hasil dari proses ini berupa nilai indeks dari fungsi keputusan yang terbesar yang menyatakan kelas dari data pengujian. Jika kelas yang dihasilkan dari proses klasifikasi pengujian sama dengan kelas data pengujian, maka pengenalan dinyatakan benar.



Gambar 6. Hasil Segementasi

Proses uji coba dengan menggunakan LS-SVM dilakukan pada data training dan pada data uji. Proses uji coba pada data training dilakukan untuk mendapatkan nilai gamma dan sigma yang paling optimal. Hasil dari proses uji coba pada data training dilakukan berkali-kali sehingga mendapatkan nilai akurasi 100%

Tabel 4. Konfusi Matriks

Aktual	Klasifikasi	
	Positif	Negatif
Positif	(59 TP)	(1 FN)
Negatif	(1 FP)	(59 TN)

Waktu Komputasi yang dibutuhkan untuk proses ekstraksi fitur untuk setiap dataset adalah rata-rata sebesar 0,85 detik, sedangkan waktu komputasi untuk uji coba pada data uji adalah rata-rata sebesar 0,006 detik. Dapat disimpulkan bahwa waktu komputasi untuk setiap dataset adalah rata-rata sebesar 0,856 detik.

### 4. KESIMPULAN

Percobaan yang dilakukan penggunaan 1 hidden layer ternyata lebih efektif berakurasi tinggi dan waktu tercepat dibandingkan menggunakan 2 hidden layer. Selanjutnya percobaan dengan 2 model jaringan syaraf tiruan, trainlm merupakan model yang sangat baik dalam hal waktu dan akurasi jika menggunakan 1 hidden layer.

Rata-rata akurasi tertinggi dalam proses klasifikasi jenis Buah mangga menggunakan epoch 5000, goal 0.0001 dan learning rate 0.1 yaitu hidden layer 1 dengan model trainlm jumlah rata-rata akurasi sebesar 89.95% dengan rata-rata waktu 6.2 detik. Sedangkan hidden layer 2 pada model traingdx diperoleh rata-rata akurasi 89.95% dengan waktu 19 detik.

**REKOMENDASI**

Ada beberapa rekomendasi yang dapat diberikan untuk kepentingan selanjutnya dalam pengembangan program maupun studi kasus yang hampir serupa, yaitu ketika menggunakan metode Backpropagation dengan model trainlm,

diperlukan spesifikasi komputer yang cukup, agar bisa waktu yang diperlukan lebih cepat.

Untuk penelitian selanjutnya bisa menambahkan jenis mangga berbeda. Pengembangan dari sisi tampilan dapat ditambah untuk memudahkan membaca grafik sehingga pengguna dapat melihat bagian grafik dari akurasi dan waktu yang dihasilkan.

**DAFTAR PUSTAKA**

- Alamuru, Susmitha, and Sanjay Jain. 2021. "Video Event Classification Using KNN Classifier with Hybrid Features." *Materials Today: Proceedings* (xxxx). <https://doi.org/10.1016/j.matpr.2021.03.154>.
- Kim, Nam Chul, and Hyun Joo So. 2018. "Directional Statistical Gabor Features for Texture Classification." *Pattern Recognition Letters* 112: 18–26. <https://doi.org/10.1016/j.patrec.2018.05.010>.
- Lee, Sue Han, Chee Seng Chan, Simon Joseph Mayo, and Paolo Remagnino. 2017. "How Deep Learning Extracts and Learns Leaf Features for Plant Classification." *Pattern Recognition* 71: 1–13.
- Malang, Negeri. 2016. "Klasifikasi Jenis Mangga Berdasarkan Bentuk Dan Tekstur Daun Menggunakan Metode K-Nearest Neighbor." 8: 247–53.
- Nnolim, U. A. 2018. "An Adaptive RGB Colour Enhancement Formulation for Logarithmic Image Processing-Based Algorithms." *Optik* 154: 192–215. <http://dx.doi.org/10.1016/j.ijleo.2017.09.102>.
- Parida, Priyadarsan, and Nilamani Bhoi. 2017. "2-D Gabor Filter Based Transition Region Extraction and Morphological Operation for Image Segmentation." *Computers and Electrical Engineering* 62: 119–34. <http://dx.doi.org/10.1016/j.compeleceng.2016.10.019>.
- Ristian, Uray. 2017. "Optimasi Ekstraksi Ciri Kontur Daun Menggunakan Interpolasi Nearest Neighbor Dan B-Spline Uray Ristian."
- Sari, Intan Permata, Bambang Hidayat, and Ratri Dwi Atmaja. 2016. "Perancangan Dan Simulasi Deteksi Penyakit Tanaman Jagung Berbasis Pengolahan Citra Digital Menggunakan Metode Color Moments Dan GLCM." : 215–20.
- Shokrzade, Amin, Mohsen Ramezani, Fardin Akhlaghian Tab, and Mahmud Abdulla Mohammad. 2021. "A Novel Extreme Learning Machine Based KNN Classification Method for Dealing with Big Data." *Expert Systems with Applications* 183(December 2019): 115293. <https://doi.org/10.1016/j.eswa.2021.115293>.
- Sumaiya, M. N., and R. Shantha Selva Kumari. 2017. "Gabor Filter Based Change Detection in SAR Images by KI Thresholding." *Optik* 130: 114–22. <http://dx.doi.org/10.1016/j.ijleo.2016.11.040>.
- Utami, Arum T R I et al. 2017. "Implementasi Metode Otsu Thresholding Untuk Segmentasi Citra Daun."
- Younesi, Ali, and Mehdi Chehel Amirani. 2017. "Gabor Filter and Texture Based Features for Palmprint Recognition." *Procedia Computer Science* 108(June): 2488–95. <http://dx.doi.org/10.1016/j.procs.2017.05.157>.
- Zhang, Rongfeng, Shizeng Lu, Hongliang Yu, and Xiaohong Wang. 2021. "Recognition Method of Cement Rotary Kiln Burning State Based on Otsu-Kmeans Flame Image Segmentation and SVM." *Optik* 243(June): 167418. <https://doi.org/10.1016/j.ijleo.2021.167418>.