

Klasifikasi Jenis Buah Kelengkeng Dengan Metode *K-Nearest Neighbor* (KNN) Berdasarkan Citra Warna Buah

Muhammad Akbar Anugrah Illahi¹, Widiyanto Tri Handoko²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Industri,
Universitas Stikubank, Semarang, Indonesia

E-mail: m.akbar0503@gmail.com¹, wthandoko@edu.unisbank.ac.id²

Abstract

In the study titled "Classification of Longan Fruit Types Using KNN Method Based on Fruit Color Images" with the use of the TensorFlow Framework, a series of system testing was conducted using various variations of longan fruit images, totaling 360 samples. The aim of this research was to classify longan fruit types based on the extraction of color features from fruit images. The test results showed the highest accuracy rate reached 98.7% and an average accuracy of 89.6% on the train and test data with an 80%:20% ratio. The developed application successfully distinguished five categories of longan fruit, namely diamond river longan, itoh longan, mata lada longan, red longan, and pingpong longan. This study used a multi-class dataset as the data source. By using the KNN method with a parameter $k=5$, the system was able to classify longan fruit images with 78% accuracy in the 80%:20% train-validation data split scenario. These findings provide a positive perspective on the potential application of the KNN method in classifying longan fruit types based on the extraction of color features from fruit images. This research makes a significant contribution to the development of automatic recognition and classification systems for longan fruit using image processing techniques

Keywords: Longan, Classification, KNN, Android

Abstrak

Dalam penelitian berjudul "Klasifikasi Jenis Buah Kelengkeng Menggunakan Metode KNN Berdasarkan Citra Warna Buah" dengan menggunakan Framework TensorFlow, dilakukan serangkaian pengujian sistem menggunakan berbagai variasi gambar jenis kelengkeng sebanyak 360 sampel. Penelitian ini bertujuan untuk mengklasifikasikan jenis buah kelengkeng berdasarkan ekstraksi fitur warna dari citra buah. Hasil pengujian menunjukkan tingkat keakuratan tertinggi mencapai 98,7% dan nilai akurasi rata-rata 89,6% pada data train dan test dengan perbandingan 80%:20%. Aplikasi yang dikembangkan berhasil membedakan lima kategori jenis kelengkeng, yaitu kelengkeng diamond river, kelengkeng itoh, kelengkeng mata lada, kelengkeng merah, dan kelengkeng pingpong. Penelitian ini menggunakan dataset multikelas sebagai sumber data. Dengan menggunakan metode KNN dan parameter $k=5$, sistem mampu mengklasifikasikan gambar jenis kelengkeng dengan akurasi 78% pada skenario perbandingan data train-validation 80%:20%. Temuan ini memberikan pandangan positif mengenai potensi penerapan metode KNN dalam klasifikasi jenis buah kelengkeng berdasarkan ekstraksi fitur warna dari citra buah. Penelitian ini memberikan kontribusi signifikan dalam pengembangan sistem pengenalan dan klasifikasi otomatis untuk buah kelengkeng menggunakan teknik pengolahan citra.

Kata Kunci: Kelengkeng, Klasifikasi, KNN, Android

1. Pendahuluan

Tanaman kelengkeng (*Dimocarpus longan* Lour) berasal dari India timur, Burma, atau Cina. Di Indonesia, terdapat dua jenis kelengkeng yang umum dibudidayakan, yaitu

kelengkeng lokal dan kelengkeng introduksi. Kelengkeng lokal memiliki beberapa kultivar, seperti kelengkeng batu, kelengkeng aroma durian, dan kelengkeng kopyor, sementara kelengkeng introduksi berasal dari Thailand (Diamond River) dan Vietnam (pingpong). Kelengkeng memiliki adaptasi yang baik dan dapat tumbuh baik di dataran rendah maupun tinggi.

Setiap jenis kelengkeng memiliki keunggulan dan kelemahan yang berbeda. Sebagai contoh, kelengkeng lokal varietas batu memiliki buah yang lebih besar, daging yang lebih tebal, dan mudah dikupas, tetapi jumlah buahnya cenderung sedikit dibandingkan dengan jenis lainnya. Buah kelengkeng memiliki bentuk bulat dan warna kulit yang berbeda-beda untuk setiap jenis. Selain itu, buah kelengkeng juga memiliki rasa manis dan kaya akan vitamin C, sehingga diminati oleh masyarakat Indonesia [1].

Namun, perbedaan ciri-ciri antara jenis kelengkeng menyulitkan dalam membedakan varietasnya. Ciri atau fitur seperti warna kulit, tekstur, bentuk, dan ukuran digunakan untuk membedakan jenis tanaman kelengkeng. Meskipun jenis-jenis kelengkeng memiliki ciri-ciri yang hampir sama, hal ini menjadi tantangan bagi orang awam atau penikmat tanaman kelengkeng dalam mengidentifikasi jenis buah kelengkeng.

Oleh karena itu, dibutuhkan sistem identifikasi otomatis menggunakan komputer untuk memudahkan pengenalan dan identifikasi jenis-jenis tanaman kelengkeng. Salah satu metode yang digunakan adalah *K-Nearest Neighbor Classifier* (KNN) dengan pengolahan citra berbasis warna kulit buah. Ekstraksi fitur warna kulit buah digunakan untuk mengenali jenis varietas kelengkeng, sehingga sistem yang dikembangkan dapat melakukan klasifikasi berdasarkan ciri-ciri tersebut.

Algoritma *K-Nearest Neighbor Classifier* (KNN) digunakan untuk mengklasifikasikan jenis buah kelengkeng berdasarkan nilai tetangga terdekatnya (K), dengan menggunakan ekstraksi warna kulit buah yang membantu dalam proses klasifikasi. Fitur warna ini mempermudah dalam pengolahan citra. Penelitian ini bertujuan untuk mengklasifikasikan jenis buah kelengkeng berdasarkan citra warnanya, dengan menggunakan nilai RGB sebagai ekstraksi fitur untuk mendapatkan ciri warna pada citra. Penelitian ini berjudul "Klasifikasi Jenis Buah Kelengkeng Dengan Metode K-Nearest Neighbor (KNN) Berdasarkan Citra Warna Buah", dan diharapkan dapat memberikan kontribusi dalam pengembangan sistem pengenalan dan klasifikasi otomatis untuk buah kelengkeng menggunakan teknik pengolahan citra.

2. Metodologi Penelitian

2.1. Kelengkeng

Buah longan (*Dimocarpus longan*) adalah buah berdaging yang berasal dari pohon longan yang tumbuh di Asia Tenggara dan China Selatan. Buah longan memiliki kulit berwarna coklat kemerahan yang kaku dan mudah dikupas. Daging buahnya transparan dan berair, dengan tekstur lembut dan rasa manis. Buah longan sering dimakan segar sebagai camilan atau digunakan sebagai bahan dalam berbagai hidangan makanan dan minuman, terutama dalam masakan Asia. Selain itu, buah longan juga mengandung nutrisi seperti vitamin C, zat besi, fosfor, dan kalium, yang bermanfaat untuk kesehatan tubuh [2].

2.2. Pengolahan Citra Digital

Pengolahan citra digital adalah bidang ilmu yang fokus pada peningkatan kualitas gambar, transformasi gambar, pemilihan ciri (feature images) optimal untuk analisis, pemrosesan informasi dalam citra, kompresi atau reduksi data untuk penyimpanan, transmisi, dan pengolahan data dengan menggunakan teknologi digital. Tujuan utama dari pengolahan citra digital adalah untuk meningkatkan kualitas citra, memahami informasi yang terkandung dalam citra, dan mempermudah manipulasi serta analisis citra menggunakan algoritma dan metode komputasi [3].

2.3. Ekstraksi Fitur Warna

Ekstraksi fitur warna RGB melibatkan pengambilan informasi mengenai komponen warna merah (*Red*), hijau (*Green*), dan biru (*Blue*) dari setiap piksel dalam sebuah citra. Dalam proses ini, intensitas warna pada ketiga komponen tersebut dijadikan sebagai fitur numerik yang merefleksikan pola warna dalam citra [4]. Tujuan utama dari ekstraksi fitur warna RGB adalah untuk menggambarkan karakteristik warna dalam citra secara terstruktur. Hasil ekstraksi fitur warna RGB dapat digunakan dalam berbagai proses analisis seperti deteksi objek, segmentasi citra, pengenalan pola, identifikasi warna, dan klasifikasi citra [5].

2.4. KNN

Algoritma *K-Nearest Neighbor* (KNN) merupakan sebuah metode non-parametrik yang digunakan untuk melakukan klasifikasi pada objek. Prinsip dasar dari algoritma KNN adalah apabila kita memiliki sekelompok sampel data sebagai data training dan memberikan label untuk seluruh data tersebut, maka peneliti dapat mengetahui kelas mana data tersebut masuk. Ketika diberikan data baru tanpa label, algoritma akan membandingkan data tersebut dengan data yang sudah ada dan mencari kesamaan untuk mencari labelnya [6].

Ada beberapa tahap dalam menghitung metode KNN sebagai berikut:

1. Peneliti menentukan nilai K (jumlah tetangga terdekat) yang akan digunakan dalam proses klasifikasi.
2. Menghitung jarak antara setiap objek dalam data sampel dan *instance* yang akan diklasifikasikan dengan menggunakan rumus *Euclidean Distance*:
3. Membaca data training dan data testing format gambar.
4. Melakukan pengujian pada 5 jenis kelengkeng dengan memasukkan nilai K.
5. Menghitung klasifikasi jenis kelengkeng menggunakan metode KNN dengan menggunakan rumus *Euclidean Distance* seperti yang dijelaskan pada tahap 2.
6. Hasil klasifikasi *instance* yang akan diklasifikasikan ditentukan berdasarkan mayoritas kategori dari tetangga terdekat.

2.5. Metode Pengumpulan Data

Dalam tahap Studi Pustaka, dilakukan pengumpulan data dan informasi yang relevan terkait dengan 5 jenis kelengkeng yang akan diklasifikasikan menggunakan metode KNN, yaitu Kelengkeng Diamond, Kelengkeng Itoh, Kelengkeng Mata lada, Kelengkeng Merah, dan Kelengkeng Pingpong. Sumber informasi yang digunakan mencakup jurnal, buku, dan artikel yang terkait dengan topik penelitian.

Observasi non-partisipatoris dilakukan oleh peneliti untuk mengumpulkan data berupa gambar-gambar jenis kelengkeng dari internet. Tahap awal meliputi pencarian gambar jenis kelengkeng dan pengelompokan gambar sesuai dengan kategori yang telah ditentukan. Data-data tersebut kemudian dikumpulkan dari internet dan dijadikan sebagai dataset untuk melatih dan menguji system.

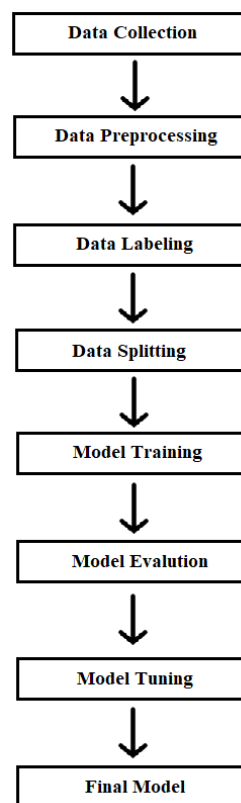
2.6. Analisis Data

Dalam penelitian ini, langkah pertama yang dilakukan adalah mengumpulkan dataset gambar kelengkeng yang nantinya akan digunakan dalam analisis. Dataset dapat diperoleh dari berbagai sumber, seperti situs web, basis data online, atau melalui pengambilan foto sendiri. Setelah data terkumpul, tahap selanjutnya adalah melakukan preprocessing data. Pada tahap ini, dilakukan normalisasi gambar, resizing gambar, dan ekstraksi fitur gambar agar dapat digunakan sebagai input dalam model. Pada tahap labeling data, setiap gambar diberi label atau kelas yang sesuai dengan jenis kelengkeng yang direpresentasikan dalam gambar yaitu kelengkeng diamond river, kelengkeng itoh, kelengkeng matalada, kelengkeng merah, kelengkeng pingpong. Label ini akan digunakan sebagai referensi dalam pembuatan model. Selanjutnya dataset dibagi menjadi data training dan data testing

dengan proporsi 80%:20%. Data training digunakan untuk melatih model dan data testing digunakan untuk menguji performa model.

Proses training model dilakukan dengan menggunakan data training yang telah diproses pada tahap sebelumnya. Pada metode KNN, model dilatih dengan menentukan parameter k (jumlah tetangga terdekat yang akan digunakan) dan melakukan perhitungan jarak antara data uji dengan data latih untuk menentukan kelas data uji. Setelah model dilatih, tahap selanjutnya adalah melakukan evaluasi model dengan menggunakan data testing yang telah diproses pada tahap sebelumnya. Performa model diukur dengan menghitung nilai akurasi, presisi, recall, dan F1-score.

Pada tahap terakhir, dilakukan tuning parameter k untuk meningkatkan performa model. Tuning parameter k dilakukan dengan mencoba beberapa nilai k dan memilih nilai k yang menghasilkan performa model terbaik. Dalam keseluruhan proses analisis data, tahap yang paling penting adalah data preprocessing dan ekstraksi fitur gambar. Kualitas gambar yang baik dan fitur yang relevan akan sangat mempengaruhi akurasi dan performa model. Selain itu, pada tahap model tuning, pemilihan parameter k yang optimal juga akan sangat berpengaruh pada performa model dan harus dilakukan secara teliti.



Gambar 1. Diagram alir metode analisis data

2.7. Klasifikasi KNN

Algoritma KNN digunakan untuk proses klasifikasi data citra. Terdapat beberapa tahapan dalam penggunaan KNN yaitu:

- Peneliti menentukan nilai K (jumlah tetangga terdekat) yang akan digunakan dalam proses klasifikasi.
- Menghitung jarak antara setiap objek dalam data sampel dan instance yang akan diklasifikasikan dengan menggunakan rumus Euclidean Distance:
- Membaca data train dan data test sebanyak 360 sampel gambar.
- Melakukan pengujian pada 5 jenis kelengkeng dengan memasukkan nilai K .

- e. Menghitung klasifikasi jenis kelengkeng menggunakan metode KNN dengan menggunakan rumus Euclidean Distance seperti yang dijelaskan pada tahap 2.
- f. Tampilan Hasil klasifikasi.

3. Hasil dan Pembahasan

3.1. Pengumpulan data

Penelitian ini melakukan klasifikasi jenis kelengkeng menjadi 5 kategori yaitu kelengkeng diamond river, kelengkeng itoh, kelengkeng matalada, kelengkeng merah, kelengkeng pingpong. Data yang digunakan terdiri dari 360 sampel gambar buah kelengkeng. Selanjutnya, dalam penelitian ini akan dilakukan ekstraksi fitur warna RGB dari setiap gambar untuk menganalisis pola warna karakteristik pada masing-masing gambar. Dari sampel tersebut, 80% digunakan untuk data training dan 20% untuk data validasi pada setiap jenis kelengkeng. Model KNN akan dilatih menggunakan data tersebut untuk menghasilkan hasil klasifikasi dengan akurasi yang tinggi.

Tabel 1. Skenario pembagian data

Jenis Kategori	Data Train 80%	Data Test 20%
Kelengkeng Diamond River	60	12
Kelengkeng Itoh	60	12
Kelengkeng Matalada	60	12
Kelengkeng Merah	60	12
Kelengkeng Pingpong	60	12
Total	300	60

3.2. Pra proses

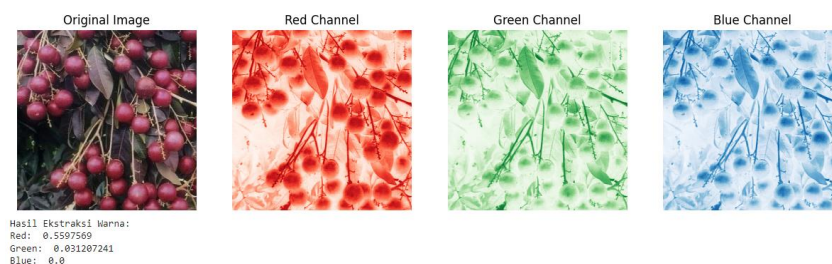
Setelah mendapatkan data, langkah selanjutnya adalah melakukan pra-pemrosesan data. Pra-pemrosesan data bertujuan untuk mendapatkan fitur-fitur dari citra kelengkeng, sehingga memudahkan dalam mengubah dataset menjadi representasi RGB. Salah satu tahapannya adalah melakukan resize citra. Pra-pemrosesan ini dilakukan untuk menyesuaikan ukuran citra agar sesuai dengan kebutuhan analisis selanjutnya.

3.3. Proses resize

Pada tahap awal praproses, dilakukan proses resize citra. Proses ini bertujuan untuk mengubah ukuran citra awal menjadi citra dengan ukuran 224x224 piksel. Proses resize ini memiliki manfaat untuk mempercepat proses pada Google Colab dan memastikan mendapatkan hasil yang sesuai dengan format yang diharapkan.

3.4. Ekstraksi fitur warna

Langkah berikutnya adalah melakukan ekstraksi fitur warna menggunakan metode RGB. Pada proses ini, nilai-nilai warna RGB dari citra akan diekstraksi. Tahap ini melibatkan pengambilan citra "10.png" yang telah melalui tahap praproses resize.



Gambar 2. Hasil ekstraksi fitur warna RGB

3.5. Hasil Ekstraksi fitur warna

Sebanyak 300 data training dikelompokkan menjadi 60 data citra kelengkeng diamond river, 60 data citra kelengkeng itoh, 60 data citra kelengkeng matalada, 60 data citra kelengkeng merah, 60 data citra kelengkeng pingpong.

Tabel 2. Hasil ekstraksi fitur warna data train

Citra Ke	R	G	B	Kelas
1	0.0959	0.0297	0.0002	Diamond
2	0.0662	0.0714	0.0020	Diamond
3	0.1320	0.0713	0.0008	Diamond
4	0.5472	0.0449	0.0019	Diamond
5	0.0249	0.0608	0.0070	Diamond
6	0.0053	0.0454	0.0001	Diamond
7	0.0193	0.0198	0.0069	Diamond
8	0.1475	0.1432	0.0121	Diamond
9	0.2887	0.0813	0.0016	Diamond
10	0.2160	0.0683	0.0023	Diamond
...	...	...	...	...
300	0.0572	0.0076	3.324	Pingpong

Data testing yang digunakan sebanyak 60 data yang dibagi menjadi 12 data citra kelengkeng diamond river, 12 data citra kelengkeng itoh, 12 data citra kelengkeng matalada, 12 data citra kelengkeng merah, 12 data citra kelengkeng pingpong.

Tabel 3. Hasil ekstraksi fitur warna data test

Citra Ke	R	G	B	Kelas
1	0.0959	0.0297	0.0002	Diamond
2	0.1475	0.1432	0.0121	Diamond
3	0.0236	0.0476	0.0025	Diamond
4	0.0476	0.0975	0.0079	Diamond
5	0.2678	0.0754	0.0019	Diamond
6	0.1191	0.0538	0.0017	Diamond
7	0.0645	0.0497	0.0013	Diamond
8	0.2630	0.0244	0.0002	Diamond
9	0.3184	0.0471	0.0171	Diamond
10	0.2923	0.0389	0.0132	Diamond
...	...	...	...	...
60	0.0572	0.0076	3.324	Pingpong

3.6. Hasil Klasifikasi KNN

Setelah melalui beberapa proses dalam algoritma K-Nearest Neighbors (KNN) didapatkan hasil training dan validation. Proses ini menggunakan k=5 dan ekstraksi fitur warna. Berikut hasil yang diperoleh berdasarkan parameter yang telah ditentukan:

Laporan Klasifikasi:				
	precision	recall	f1-score	support
diamond	0.70	0.58	0.64	12
itoh	0.59	0.83	0.69	12
matalada	0.80	1.00	0.89	12
merah	1.00	0.92	0.96	12
pingpong	1.00	0.58	0.74	12
accuracy			0.78	60
macro avg	0.82	0.78	0.78	60
weighted avg	0.82	0.78	0.78	60

Gambar 3. Classification Report

Gambar.3 memperlihatkan hasil laporan klasifikasi menunjukkan performa model K-Nearest Neighbors (KNN) dengan parameter $k=5$ dan ekstraksi fitur warna yang telah ditentukan. Dalam hal ini, peneliti berhasil mencapai akurasi sebesar 78% pada data validasi. Selain itu, laporan juga memberikan informasi tentang precision, recall, dan f1-score untuk setiap kategori yang diklasifikasikan.

3.7. Pengujian Model Aplikasi

Hasil tampilan di bawah ini adalah hasil dari pengujian aplikasi yang telah dilakukan dengan menggunakan berbagai data uji yang acak. Dalam pengujian ini, aplikasi dijalankan dengan berbagai input data untuk menguji fungsionalitas dan kinerja dari model yang telah dikembangkan.



Gambar 4. Klasifikasi model aplikasi

Hasil pengujian acak klasifikasi jenis kelengkeng berbasis Android dapat dilihat pada Gambar 4. Gambar tersebut menyajikan jenis kelengkeng yang diklasifikasikan dan

persentase tingkat keakuratan yang diperoleh. Dalam pengujian ini, kelengkeng diamond mencapai tingkat akurasi sebesar 84,9% (a), kelengkeng itoh mencapai tingkat akurasi sebesar 86,1% (b), kelengkeng mata lada mendapatkan tingkat akurasi sebesar 97,2% (c), kelengkeng merah mendapatkan tingkat akurasi sebesar 98,7% (d), dan kelengkeng pingpong mencapai tingkat akurasi sebesar 94,1% (e).

4. Kesimpulan

Berdasarkan penelitian yang berjudul "Klasifikasi Jenis Buah Kelengkeng Menggunakan Metode KNN Berdasarkan Citra Warna Buah" dengan menggunakan Framework Tensorflow, dilakukan pengujian sistem menggunakan berbagai macam gambar jenis kelengkeng yang menghasilkan nilai yang beragam. Hasil pengujian menunjukkan bahwa sistem mencapai nilai ketepatan tertinggi sebesar 98,7% dan nilai akurasi rata-rata sebesar 89,6%. Aplikasi yang dikembangkan berhasil membedakan lima kategori jenis kelengkeng, yaitu kelengkeng diamond river, kelengkeng itoh, kelengkeng mata lada, kelengkeng merah, dan kelengkeng pingpong. Penelitian ini menggunakan dataset multikelas sebagai sumber data. Hasil akhir menunjukkan bahwa menggunakan metode KNN dengan parameter $k=5$, sistem berhasil mengklasifikasikan gambar jenis kelengkeng dengan akurasi sebesar 78% menggunakan skenario perbandingan data train validation 80%:20%.

Daftar Pustaka

- [1] Tyas, P. S., & Setyati, D. (2013). Perkembangan Pembungaan Lengkeng (*Dimocarpus longan* Lour) 'Diamond river' Flowering Development of *Longan* (*Dimocarpus longan* Lour) 'Diamond river'. *Jurnal Ilmu Dasar*, 14(2), 111-120.
- [2] Saraswati, R., Susilowati, M. D., Restuti, R. C., & Pamungkas, F. D. Departemen Geografi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia 2019.
- [3] Sutoyo, T., Si, S., & Kom, M. (2009). Teori Pengantar Citra Digital. ANDI Yogyakarta.
- [4] Wiharja, Y. P., & Harjoko, A. (2014). Pemrosesan citra digital untuk klasifikasi mutu buah pisang menggunakan jaringan saraf tiruan. *Ijeis*, 4(1), 57-68.
- [5] Setiawan, E. B., & Herdianto, R. (2018). Penggunaan Smartphone Android sebagai Alat Analisis Kebutuhan Kandungan Nitrogen pada Tanaman Padi. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 7(3), 273-280.
- [6] Budiharto, W. (2016). Machine learning & computational intelligence. ANDI, Yogyakarta.