

Klasifikasi Jenis Bunga Menggunakan Algoritma K-NN Dengan Ekstraksi Fitur HSV Dan GLCM

Dewa^{1*}, Aldi Ardiansyah², Mecha Bella Permata Sihombing³, M. Kurniawan⁴, Ery Hartati⁵

¹²³⁴Program Studi Informatika, Universitas Multi Data Palembang, Palembang, Indonesia
dewa_2226250101@mhs.mdp.ac.id, aldiardiansyah_2226250059@mhs.mdp.ac.id,
mechabellapermatasihombing_2226250095@mhs.mdp.ac.id,
mkurniawan_2226250069@mhs.mdp.ac.id, ery_hartati@mdp.ac.id

Abstrak

Beragamnya jenis bunga di alam dengan kemiripan visual yang tinggi sering menyulitkan proses identifikasi, terutama bagi masyarakat awam maupun pelaku budidaya. Proses pengenalan manual cenderung memakan waktu dan tidak akurat. Oleh karena itu, penelitian ini mengusulkan pendekatan berbasis pengolahan citra digital dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN) untuk mengklasifikasikan jenis bunga. Ekstraksi fitur dilakukan dengan metode HSV untuk warna dan GLCM untuk tekstur, sehingga citra bunga dapat direpresentasikan secara lebih menyeluruh. Dataset yang digunakan diperoleh dari situs Kaggle, terdiri dari tujuh kelas bunga dengan rasio pembagian data latih dan uji sebesar 60:40. Pengujian dilakukan pada nilai $K=3$, $K=5$, dan $K=7$, dengan akurasi tertinggi sebesar 67,86% untuk $K=3$ dan $K=5$. Hasil menunjukkan bahwa kombinasi fitur HSV dan GLCM mampu mendukung proses klasifikasi dengan hasil yang cukup baik, meskipun akurasinya masih dapat ditingkatkan dengan jumlah data lebih besar dan pengembangan metode. Penelitian ini diharapkan dapat menjadi dasar bagi pengembangan sistem klasifikasi objek serupa di bidang botani maupun pengenalan visual berbasis citra lainnya.

Kata kunci : *Klasifikasi, HSV, GLCM, K-NN, Citra Bunga*

Abstract

The variety of flowers in nature with high visual similarity often makes the identification process difficult, especially for ordinary people and cultivators. Manual recognition processes tend to be time-consuming and inaccurate. Therefore, this research proposes a digital image processing-based approach using the K-Nearest Neighbor (K-NN) algorithm to classify flower types. Feature extraction is performed using HSV method for color and GLCM for texture, so that the flower image can be represented more thoroughly. The dataset used is obtained from the Kaggle website, consisting of seven flower classes with a training and test data division ratio of 60:40. Tests were conducted on the values of $K = 3$, $K = 5$, and $K = 7$, with the highest accuracy of 67.86% for $K = 3$ and $K = 5$. The results show that the combination of HSV and GLCM features is able to support the classification process with fairly good results, although the accuracy can still be improved with a larger amount of data and method development. This research is expected to be the basis for the development of similar object classification systems in botany and other image-based visual recognition.

Keywords: *Classification, HSV, GLCM, K-NN, Flower Image*

1. PENDAHULUAN

Tumbuhan adalah makhluk hidup yang tumbuh di tanah atau media lain yang menyediakan berbagai kebutuhan seperti bahan makanan sampai menjadi penghias untuk mempercantik taman atau rumah. Salah satu tumbuhan yang menjadi favorit banyak orang adalah bunga. Bunga merupakan salah bagian dari sebuah tumbuhan yang memiliki warna dan bentuk sangat beragam dan indah [1]. Bunga mampu menghadirkan perasaan bahagia, damai, dan ketenangan. Selain warnanya yang memukau, bunga juga memancarkan aroma yang harum dan menyegarkan, serta memiliki keunikan tersendiri yang dapat diamati dari berbagai ciri khasnya [2].

Namun dengan banyaknya jenis bunga yang ada, banyak orang yang kesulitan dalam mengenali dan mengingat nama setiap bunga yang hanya dilihat sekilas. Banyak juga orang yang tidak mengenali atau sama sekali tidak mengetahui jenis-jenis bunga yang mereka lihat sehari-hari. Hal ini sangat sering terjadi karena mengingat banyaknya variasi dan keunikan bunga yang ada, sehingga untuk mengenalinya menjadi sebuah tantangan bagi banyak orang.

Salah satu cara untuk mengetahui jenis dan ciri yang beragam dari bunga adalah melalui proses klasifikasi. Klasifikasi adalah pengelompokan sesuatu berdasarkan kesamaan atau ciri tertentu agar mudah untuk dipahami. Proses pengelompokan ini bisa dilakukan dengan berbagai cara, baik secara manual maupun dengan bantuan teknologi. Proses klasifikasi biasanya memanfaatkan berbagai algoritma cerdas salah satu nya adalah *K-Nearest Neighbor* [3].

Metode yang efektif dalam mengklasifikasi jenis dan ciri bunga adalah algoritma *K-Nearest Neighbor* (KNN). Algoritma KNN merupakan metode pembelajaran mesin yang sering digunakan untuk mengklasifikasi data berdasarkan kemiripan dengan data lain yang sudah diketahui. Menurut penelitian yang dilakukan oleh [4]

pada tahun 2022 tentang klasifikasi jenis bunga mawar menggunakan metode KNN, hasilnya menunjukkan bahwa KNN memiliki akurasi sebesar 92%. Penelitian yang dilakukan oleh [5] pada tahun 2017 mengenai klasifikasi jenis pisang menggunakan metode KNN memperoleh akurasi 80%. Dan penelitian yang dilakukan oleh [6] tentang klasifikasi kematangan buah pisang ambon menggunakan metode KNN menunjukkan hasil akurasi saat $K=5$ sebesar 90.9%.

Berdasarkan permasalahan diatas penelitian ini bertujuan untuk mengklasifikasi jenis bunga, berdasarkan dataset yang diambil dari kaggle. Penelitian ini menggunakan algoritma KNN yang digunakan untuk mengklasifikasi data seperti gambar. Diharapkan hasil penelitian ini dapat memberikan wawasan bagi orang-orang yang belum mengetahui berbagai jenis bunga.

2. METODE

Metodologi Pada Laporan ini berdasarkan pada tahapan pengolahan citra digital yang dilakukan meliputi seperti dibawah ini :

2.1. Data

Data Pada laporan proyek ini menggunakan Data public yang diambil langsung diambil dari website kaggle yang berupa terdapat dua macam Data Jenis bunga yaitu :

- 1) Data Latih (Data Training), Yang dimana digunakan sebagai data latih untuk model K-NN terdiri dari 7 kelas atau 7 jenis bunga yang masing-masing jenis sebanyak 10 citra
- 2) Data Uji (Testing Data), Merupakan data yang digunakan untuk menguji model K-NN dimana mengukur kinerjanya setelah dilakukan proses pelatihan yang sudah diberikan sebelumnya.

Berikut Merupakan beberapa Gambar dari jenis atau kelas bunga yang kami gunakan untuk dilakukan proses klasifikasi yang

dimana kami mengambilnya langsung dari website kaggle yaitu sebagai berikut:



(1)



(2)



(3)



(4)



(5)



(6)



(7)

Gambar 1. (1) Bunga Rose (2) Bunga Calendula (3) Bunga Iris (4) Bunga Dandelion (5) Bunga Viola (6) Bunga Lilium (7) Bunga Water Lily

2.2. Algoritma

Dalam Proses untuk Mengklasifikasi kami menggunakan algoritma K-NN sebagai

alternatif karena beberapa alasannya, Algoritma ini cukup banyak digunakan untuk mengklasifikasi suatu citra terkhusus yang kami gunakan sekarang untuk mengklasifikasikan jenis bunga yang terdiri 7 jenis bunga tersebut. Algoritma K-NN disini bekerja dengan beberapa tahapan yaitu sebagai berikut:

- 1) Menghitung jarak atau distance misalnya seperti pada proyek akhir ini kami menggunakan jarak atau distance Euclidean yang dimana menghitung jarak antara data uji dengan semua data latih.
- 2) Menentukan k tetangga terdekat (*nearest neighbors*) yang dimana dalam prosesnya, untuk pemilihan nilai K cukup krusial karena cukup mempengaruhi dalam performa klasifikasi maka disini kami menguji antara $k = 3$, $k = 5$, dan $k = 7$ yang dimana kami membandingkan nilai K-tetangga mana yang terbaik untuk digunakan dalam mengklasifikasi Jenis bunga.
- 3) Menentukan label kelas berdasarkan mayoritas dari k tetangga tersebut seperti pada proyek akhir ini setiap kelas diberi label sesuai dengan nama jenis bunga tersebut sehingga memudahkan dalam mengklasifikasi jenis bunga tersebut.

2.3. Rancangan Sistem

Dalam Perancangan sistem disini dilakukan dalam beberapa proses seperti di bawah ini :

1) Pengumpulan data :

Tujuan pada tahap ini adalah untuk menentukan data yang dibutuhkan dan pemilihan metode pengolahan citra digital. Langkah-langkah yang dilakukan dalam ini secara umum dimulai dari persiapan objek yang akan diambil citranya, sampai proses pencitraan . Pada proses ini pengambilan citra dilakukan di <https://www.kaggle.com/> dengan judul Flower Color Image data citra akan

digunakan sebagai sampel untuk identifikasi berdasarkan warna dan tekstur pada jenis bunga. Data citra hanya menggunakan jenis bunga.

2) *Pre Processing* :

Proses pre processing digunakan untuk memperoleh citra yang akan dilanjutkan dan dimanfaatkan untuk sistem yang dirancang. Proses ini terdiri dari proses segmentasi dan menghilangkan noise pada citra. Setiap citra jenis bunga memiliki 10 dengan format .png. Lalu dilakukannya segmentasi untuk mengubah citra jenis bunga RGB (Red, Green, Blue) menjadi *citra grayscale* terlebih dahulu dengan *otsu thresholding*. Lalu dilanjutkan dengan operasi morfologi sebagai penghilang noise.

1. **Cropping** : Proses cropping adalah proses memotong bagian tertentu dari sebuah gambar atau citra digital untuk mendapatkan area yang benar-benar dibutuhkan. Proses ini bertujuan agar citra yang diolah menjadi lebih fokus dan mudah di analisis. Dalam tahap ini, ditentukan 2 titik penting yaitu titik awal dan titik akhir. Titik awal menunjukkan dari mana citra mulai dipotong, sedangkan titik akhir menandai batas potongannya[7].
2. **Segmentasi** : Segmentasi citra adalah proses dalam pengolahan gambar yang bertujuan memisahkan bagian objek dari latar belakangnya. Dengan memisahkan kedua area tersebut, objek yang akan dianalisis menjadi lebih jelas dan mudah dikenali. Proses ini sangat bergantung pada kemampuan visual untuk membedakan antara objek dan sekitarnya agar analisis dapat dilakukan dengan tepat [8].
3. **Otsu Thresholding** : Otsu thresholding adalah salah satu cara yang digunakan untuk memisahkan bagian-bagian dalam citra digital secara otomatis. Metode ini

mengubah gambar berwarna abu-abu menjadi hitam putih dengan menentukan batas (threshold) yang tepat berdasarkan perbandingan antara nilai batas tersebut dan warna setiap titik dalam gambar [9].

4. **Morfologi** : Morfologi digunakan sebagai langkah pertama maupun terakhir dari proses analisis citra. Morfologi digunakan untuk mengekstrak bagian-bagian penting dari gambar yang membantu dalam menggambarkan dan mendeskripsikan suatu area [10]. Metode yang digunakan dalam morfologi proyek ini adalah operasi *Gradient*. Morfologi *Gradient* adalah perbedaan antara dilation dan erosion dari suatu gambar. operasi morfologi juga berguna untuk menghilangkan noise.

3) *Ekstraksi Ciri* :

Warna HSV : HSV digunakan sebagai ekstraksi ciri dengan seleksi warna berdasarkan nilai Hue, Saturation, dan Value. Hue membantu membedakan berbagai warna dan menentukan seberapa merah, hijau, atau warna lainnya yang tampak dalam cahaya tersebut, karena terkait dengan panjang gelombang cahaya. Saturation menunjukkan seberapa murni warna itu, atau seberapa sedikit warna putih yang tercampur di dalamnya. Dan value adalah tingkat kecerahan warna, yaitu seberapa banyak cahaya yang diterima tanpa mempedulikan jenisnya [11]. Proses untuk mendapatkan nilai dari setiap warna yang ingin ditampilkan melalui proses perhitungan dengan melakukan konversi ruang warna RGB ke ruang warna HSV.

Tekstur GLCM : GLCM digunakan untuk mengamati pola tekstur dalam sebuah gambar dengan cara menghitung berbagai fitur seperti kontras, homogenitas, energi, dan entropi. Fitur-fitur ini membantu memberikan gambaran tentang struktur lokal dalam gambar tersebut. Metode ini banyak digunakan di berbagai bidang, seperti pengenalan pola, klasifikasi gambar, bahkan

dalam dunia medis untuk mendeteksi pola-pola tidak biasa pada gambar medis. Dengan begitu, GLCM menjadi alat yang sangat berguna untuk memahami bagaimana piksel-piksel dalam gambar saling berhubungan dan membantu analisis citra secara lebih mendalam[12].

4) K-Nearest Neighbor :

KNN (*K-Nearest Neighbor*) adalah metode klasifikasi objek berdasarkan jarak terdekat dengan objek atau fitur, data yang digunakan paling umum dalam data pembelajaran. Data pelatihan diproyeksikan ke dalam ruang multidimensi, di mana setiap dimensi memiliki fitur pada data. Ruang dibagi menjadi beberapa bagian sesuai dengan klasifikasi data pelatihan. Sebuah titik dalam ruang ini ditandai sebagai kategori c, kategori c adalah kategori yang paling umum di antara k tetangga terdekat titik titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak euclidean dengan rumus sebagai berikut :

$$D_{ij} = \sqrt{\sum_{K=1}^n (i - j)^2}$$

Keterangan:

D_{ij} : jarak Euclidean antara i dan j.

i : data pada x ke-i untuk tahap perhitungan.

j : data pada y ke-j untuk tahap perhitungan.

Klasifikasi *K-Nearest Neighbors*(KNN) bekerja dengan cara mencari sejumlah tetangga terdekat dari data yang ingin diuji, kemudian menentukan kelas berdasarkan mayoritas tetangga tersebut. Berikut langkah-langkah dalam melakukan klasifikasi menggunakan KNN :

- 1) Menentukan nilai K, yaitu jumlah tetangga terdekat yang akan di perhitungkan.
- 2) Hitung jarak Euclidean kuadrat antara data uji dengan setiap data latih yang ada.
- 3) Urutkan data latih berdasarkan jarak terkecil ke data uji.
- 4) Kumpulkan kategori atau kelas dari k tetangga terdekat tersebut.
- 5) Pilih kelas yang paling banyak muncul

diantara tetangga tersebut sebagai prediksi untuk data uji.

K-Nearest Neighbors(KNN) punya beberapa keunggulan, seperti kemampuannya yang cukup tahan terhadap data latih yang mengandung banyak noise, serta bekerja efektif ketika jumlah data latihnya besar. Namun, metode ini juga memiliki beberapa kelemahan. Salah satunya adalah kita harus menentukan nilai parameter k terlebih dahulu, yaitu jumlah tetangga terdekat yang akan diperhitungkan. Selain itu, KNN bergantung pada pengukuran jarak, tapi kadang tidak jelas jenis jarak mana yang paling tepat digunakan, serta atribut mana yang harus dipilih agar hasilnya optimal. Di sisi lain, KNN membutuhkan waktu komputasi yang cukup besar karena harus menghitung jarak antara setiap data uji dengan seluruh data latih yang ada[13].

3. HASIL DAN PEMBAHASAN

Metodologi Pada Laporan ini berdasarkan pada tahapan pengolahan citra digital yang dilakukan meliputi seperti dibawah ini :

3.1. Pengujian

Pada tahap ini dilakukan pengujian hasil dari klasifikasi dengan perhitungan probabilitas kemiripan data lama (Data Pelatihan) pelatihan yang telah diberikan untuk model pembelajaran model K-NN dengan data yang baru (Data Uji). Pada tahapan ini akan dilakukan proses pengujian perhitungan akurasi untuk setiap nilai K, yang diantaranya adalah K = 3, K = 5, K = 7 dimana apakah nilai K tetangga tersebut berpengaruh terhadap akurasi hasil dari klasifikasi jenis bunga sehingga dapat membuktikan nilai K hasil mana yang lebih baik untuk digunakan dalam proses klasifikasi ini. Berikut merupakan hasil hasil pengujian yang telah kami lakukan.

Tabel 1. Hasil Pengujian

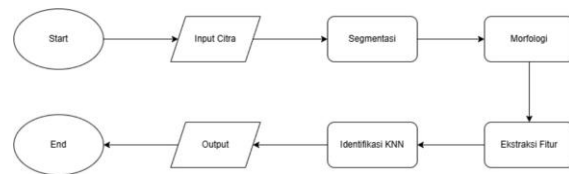
No	Split Data	Nilai K	Akurasi
1	60% Latih 40% Data Uji	K = 3	67.86%
2		K = 5	67.86%
3		K = 7	64.29%

Berdasarkan pengujian yang dilakukan yaitu dengan pembagian data menggunakan 60% data latih dan 40% data uji dari total keseluruhan 70 citra dengan 10 citra per kelas atau jenis bunga dibagi menjadi dua data yaitu 70 untuk data latih dan 28 data uji, didapatkan hasil akurasi identifikasi berbeda-beda dengan K-3 = 67.86%, K-5 = 67.86% dan K-7 = 64.29%. Pengujian dengan nilai K dan pembagian data ini dilakukan untuk mengetahui pengaruh rasio atau persentase tertentu dari jumlah data training dan data testing terhadap tingkat akurasi algoritma KNN yang digunakan. Setelah proses pengujian berhasil dilakukan disini kami selanjutnya memilih akurasi terbaik dan nilai K terbaik yang digunakan untuk identifikasi pada sebuah GUI di MATLAB.

3.2. Analisis

Berdasarkan beberapa proses yang telah dilakukan maka hasil Identifikasi Citra menggunakan algoritma K-NN dengan ekstraksi fitur HSV dan GLCM untuk jenis bunga berhasil dilakukan dengan pembagian data 60:40 serta pengujian pada nilai K-3, K-5, K-7. Pada pengujian pembagian data 60:40 dengan menggunakan nilai K-3 dan K-5 yang dimana mendapatkan nilai akurasi terbaik yaitu sebesar 67.86%, hasil ini yang didapatkan dirasa kurang dikarenakan data yang didapat sangat terbatas untuk digunakan sehingga mungkin membutuhkan data pelatihan yang lebih lagi dan fitur

ekstraksi tambahan. Diharapkan juga pada kegiatan proyek selanjutnya bisa menambahkan jumlah inputan citra jenis bunga agar mendapatkan akurasi yang lebih baik serta penerapan algoritma dan ekstraksi ciri lain yang lebih mendukung untuk mengidentifikasi jenis bunga tersebut. Berikut ini merupakan gambar dari alur kerja dari proses klasifikasi dalam mengidentifikasi jenis bunga menggunakan algoritma K-NN dengan fitur ekstraksi GLCM dan HSV yang telah dilakukan :



Gambar 2. Alur Proses Klasifikasi

4. KESIMPULAN

Kesimpulan yang didapatkan pada kegiatan untuk proses klasifikasi untuk mengidentifikasi Citra Jenis Bunga menggunakan algoritma K-NN dengan fitur ekstraksi Warna HSV dan Tekstur GLCM berhasil untuk dilakukan. Dimana dengan pembagian data 60:40 serta pengujian pada nilai K=3, K=5, K=7 yang disini mendapatkan akurasi terbaik pada pembagian data 60:40 dengan menggunakan nilai K=7 dan K=5 sebesar 67.86%, hasil yang didapatkan dirasa kurang dikarenakan data yang didapat sangat terbatas untuk digunakan. Performa untuk mengklasifikasi jenis bunga disini juga dapat dipengaruhi oleh kualitas gambar atau citra data uji yang diinputkan, seperti pencahayaan dan noise.

DAFTAR PUSTAKA

- [1] Farokhah, L. (2020). Implementasi K-Nearest Neighbor untuk klasifikasi bunga dengan ekstraksi fitur warna RGB. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), 7(6), 1129–1136. <https://doi.org/10.25126/jtik.202072608>.

- [2] Worung, D. T., Sompie, S. R. U. A., & Jacobus, A. (2021). Implementasi K-Means dan K-NN pada pengklasifikasian citra bunga. *Technomedia Journal (TMJ)*, 6(1), 124–137.
<https://ejournal.unsrat.ac.id/index.php/informatika/article/download/31965/30363>.
- [3] Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2022). Metode-metode klasifikasi. *Jurnal SAKTI*, 1(2), 15–26.
<https://e-journals.unmul.ac.id/index.php/SAKTI/article/view/2101>.
- [4] Hidayat, T., Handayani, Y., & Syaifudin, A. (2023). Implementasi algoritma K-Nearest Neighbor untuk meningkatkan penjualan produk meuble dan furniture. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, 9(2), 118–124.
<https://ejournal.uin-suska.ac.id/index.php/RMSI/article/download/24948/9630>.
- [5] Prayogo, S., Chamid, A. A., & Murti, A. C. (2022). Perancangan sistem klasifikasi jenis bunga mawar menggunakan metode K-Nearest Neighbor (KNN). *Indonesian Journal of Technology, Informatics and Science (IJTIS)*, 3(2), 52–56.
<https://jurnal.umk.ac.id/index.php/ijtis/article/view/7881/pdf>.
- [6] Iklima, C. P., Nasir, M., & Hidayat, H. T. (2017). Klasifikasi jenis pisang menggunakan metode K-Nearest Neighbor (KNN). *Jurnal Teknologi Rekayasa Informasi dan Komputer*, 1(1), 1-7. <http://ejournal.pnl.ac.id/TRIK/article/view/1854/1623>.
- [7] Wulandari, R. D., & Falah, F. (2022). Identifikasi karakteristik morfologi dan tekstur tanaman selada dengan segmentasi HSV dan GLCM. *Jurnal SINUS*, 6(2), 93–101.
https://p3m.sinus.ac.id/jurnal/index.php/e-jurnal_SINUS/article/view/664/pdf.
- [8] Pradana, A., & Putra, A. E. (2014). Pengenalan pola citra digital menggunakan metode jaringan syaraf tiruan (JST). *Seminar Nasional Informatika (SEMNASIF)*.
<https://media.neliti.com/media/publications/107084-ID-none.pdf>.
- [9] Supriyanto, A., & Sundari, R. (2021). Penerapan algoritma Otsu Thresholding dalam segmentasi citra digital. *Journal of System and Information Technology (JoSYC)*, 1(2), 35–41.
<https://ejurnal.seminarid.com/index.php/josyc/article/download/399/343/>.
- [10] Sanjaya, R. (2024, 28 januari). Pertemuan 5: Morfologi citra [Presentasi]. Scribd.
<https://www.scribd.com/document/701893197/Pertemuan-5-Morfologi-Citra>.
- [11] Suryawan, L. (2018). Segmentasi citra menggunakan metode Otsu Thresholding. *Jurnal Informatika*, 3(1), 22–30.
<https://journal.ukrim.ac.id/index.php/JIF/article/view/304/241>.
- [12] Zahrotun, N., & Widodo, I. (2022). Penerapan metode Gray Level Co-occurrence Matrix (GLCM) dalam mengenali citra daun jambu biji. *Jurnal Teknologi dan Terapan (JTT)*, 3(2), 149–156.
<https://journal.nurulfikri.ac.id/index.php/JTT/article/view/1451/381>.
- [13] Saputra, D. P., & Sarwono, J. (2022). Analisis perbandingan ekstraksi ciri warna RGB dan HSV menggunakan metode K-Nearest Neighbor (K-NN). *Jurnal SKANIKA*, 6(2), 127–135.
<https://jom.fti.budiluhur.ac.id/index.php/SKANIKA/article/view/2951>.