

Analisis Pengaruh Kondisi Akuisisi Citra terhadap Klasifikasi Butterfly Fish Menggunakan Fitur HSV dan KNN

I Putu Arya Putra, Wayan Eka Ariawan

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Tabanan, Indonesia
tuaryaputra@gmail.com, ekaariawan42@gmail.com

Info Artikel

Riwayat Artikel:

Received 2025-12-24

Revised 2026-05-01

Accepted 2025-05-14

Abstract – Image acquisition conditions are a crucial factor in fish image classification; however, they are often overlooked in dataset development. Most previous studies rely on images captured under controlled backgrounds and lighting, resulting in limited evaluation of model performance in more complex real-world environments. This study aims to analyze the effect of image acquisition conditions on the classification performance of butterfly fish using HSV color features and the K-Nearest Neighbor (KNN) algorithm. The dataset is divided into two types: a conditioned dataset captured in a mini-studio with controlled lighting and background, and an unconditioned dataset obtained in an aquarium environment with varying lighting and backgrounds. Feature extraction is performed by computing the average values of Hue, Saturation, and Value, and classification is conducted using KNN with 5-fold and 10-fold cross-validation. The results show that although the conditioned dataset provides more consistent image representations, the unconditioned dataset, which has a larger quantity and higher diversity of data, achieves better classification performance. These findings indicate that while image acquisition conditions remain important for maintaining data quality, in the KNN approach, data quantity and diversity have a more dominant influence on model performance. This study suggests that dataset development for real-world applications should emphasize data diversity and quantity rather than relying solely on controlled acquisition conditions.

Keywords: Butterfly fish, conditioned dataset, unconditioned dataset, HSV, K-Nearest NeighborFish

Corresponding Author:

I Putu Arya Putra

Email:

tuaryaputra@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak – Kondisi akuisisi citra merupakan faktor penting dalam klasifikasi citra ikan, namun sering kali diabaikan dalam pengembangan dataset. Sebagian besar penelitian sebelumnya menggunakan citra dengan latar belakang dan pencahayaan terkontrol, sehingga performa model pada kondisi nyata yang lebih kompleks belum dievaluasi secara memadai. Penelitian ini bertujuan untuk menganalisis pengaruh kondisi akuisisi citra terhadap kinerja klasifikasi butterfly fish menggunakan fitur warna HSV dan algoritma K-Nearest Neighbor (KNN). Dataset dibagi menjadi dua jenis, yaitu dataset terkondisi yang diambil dalam mini studio dengan pencahayaan dan latar belakang terkontrol, serta dataset tidak terkondisi yang diperoleh dari lingkungan akuarium dengan variasi pencahayaan dan latar belakang. Ekstraksi fitur dilakukan dengan menghitung nilai rata-rata Hue, Saturation, dan Value, kemudian klasifikasi dilakukan menggunakan KNN dengan skema evaluasi 5-fold dan 10-fold cross-validation. Hasil penelitian menunjukkan bahwa meskipun dataset terkondisi menghasilkan representasi citra yang lebih konsisten, dataset tidak terkondisi yang memiliki jumlah dan keragaman data yang lebih besar mampu menghasilkan kinerja klasifikasi yang lebih baik. Temuan ini menunjukkan bahwa kondisi akuisisi citra tetap berperan penting dalam menjaga kualitas data, namun dalam pendekatan KNN, faktor jumlah dan keragaman data memiliki pengaruh yang lebih dominan terhadap performa model. Hasil ini mengindikasikan bahwa pengembangan dataset untuk aplikasi nyata sebaiknya lebih menekankan pada keberagaman dan jumlah data dibandingkan hanya pada pengendalian kondisi akuisisi.

Kata Kunci: Butterfly Fish, Dataset Terkondisi, Dataset Tidak Terkondisi, HSV, K-Nearest Neighbor.

I. PENDAHULUAN

Butterfly fish merupakan kelompok ikan karang yang memiliki pola warna khas dan nilai ekonomi yang tinggi dalam industri ikan hias. Identifikasi spesies butterfly fish menjadi kebutuhan penting untuk mendukung kegiatan industri ikan hias [1] Namun, proses identifikasi secara manual masih sangat bergantung pada keahlian pakar, bersifat subjektif, dan tidak efisien untuk skala besar.

Perkembangan pengolahan citra digital dan *machine learning* telah mendorong pemanfaatan citra sebagai dasar identifikasi spesies ikan. Berbagai penelitian terdahulu melaporkan bahwa fitur warna pada ruang warna HSV banyak dimanfaatkan dalam analisis citra ikan, terutama untuk mengkarakterisasi perubahan visual yang berkaitan dengan tingkat kesegaran ikan [2], sementara algoritma *K-Nearest Neighbor* (KNN) masih banyak digunakan karena

kesederhanaan dan kemampuannya dalam mengklasifikasikan data berbasis jarak [3]. Beberapa studi terkini menunjukkan bahwa kombinasi ekstraksi fitur warna HSV dan algoritma *K-Nearest Neighbor* (KNN) masih menjadi pendekatan yang efektif dalam berbagai domain pengolahan citra. Pada penelitian klasifikasi tingkat kematangan buah tomat, metode HSV yang dipadukan dengan KNN berhasil mencapai akurasi hingga 92.5%, terutama ketika menggunakan citra beresolusi tinggi, sehingga menegaskan performa metode ini sangat dipengaruhi oleh karakteristik dataset yang digunakan [4]. Studi lain pada klasifikasi ikan cupang dengan jumlah data terbatas (100 citra dari 10 jenis ikan) menunjukkan bahwa KNN tetap mampu memberikan hasil yang layak, meskipun akurasi hanya sekitar 70%, yang mengindikasikan bahwa algoritma sederhana ini masih relevan untuk dataset berskala kecil [5].

Sebagian besar penelitian sebelumnya berfokus pada peningkatan kinerja klasifikasi melalui optimalisasi algoritma, penerapan pra-pemrosesan citra, maupun penggunaan teknik penyeimbangan data seperti *Synthetic Minority Over-sampling Technique* (SMOTE). Pendekatan tersebut terbukti efektif dalam meningkatkan nilai akurasi [6], namun cenderung mengaburkan pengaruh karakteristik alami dataset terhadap performa model. Berbeda dengan pendekatan tersebut, penelitian ini tidak bertujuan untuk meningkatkan akurasi melalui rekayasa data atau model, melainkan difokuskan pada analisis karakteristik dataset berdasarkan kondisi akuisisi citra.

Selain itu, penelitian yang secara eksplisit membandingkan kinerja model pada dataset dengan kondisi akuisisi citra yang berbeda yaitu terkondisi (*controlled*) dan tidak terkondisi (*uncontrolled*) masih sangat terbatas. Perbedaan karakteristik dataset, seperti jumlah data yang berbeda dan keragaman visual, berpotensi memberikan pengaruh signifikan terhadap performa model, khususnya pada algoritma berbasis jarak seperti KNN yang sangat bergantung pada distribusi data dalam ruang fitur.

Berdasarkan celah penelitian tersebut, penelitian ini bertujuan untuk menganalisis pengaruh kondisi akuisisi citra terhadap kinerja klasifikasi butterfly fish menggunakan fitur HSV dan algoritma KNN. Penelitian ini secara khusus membandingkan dataset terkondisi dan tidak terkondisi dengan tetap mempertahankan karakteristik alami masing-masing dataset sebagai representasi kondisi nyata. Dengan demikian, analisis tidak hanya berfokus pada kondisi akuisisi citra, tetapi juga mempertimbangkan pengaruh jumlah dan keragaman data terhadap performa model.

Kontribusi utama penelitian ini adalah (1) Menyediakan analisis komparatif antara dataset terkondisi dan tidak terkondisi menggunakan metode yang sama. (2) Menunjukkan bahwa jumlah dan keragaman data memiliki pengaruh yang lebih dominan dibandingkan kondisi akuisisi citra dalam pendekatan KNN. (3) memberikan implikasi praktis bahwa pengembangan dataset untuk aplikasi nyata sebaiknya lebih menekankan pada keberagaman dan jumlah data dibandingkan hanya pada pengendalian kondisi akuisisi

II. METODE

A. Penelitian Terdahulu

Berbagai penelitian terdahulu telah mengeksplorasi pemanfaatan ekstraksi fitur warna dan algoritma *K-Nearest Neighbor* (KNN) dalam berbagai domain pengolahan citra. Fitur warna dalam ruang Hue, Saturation, dan Value (HSV) terbukti efektif dalam merepresentasikan karakteristik visual objek karena mampu memisahkan informasi warna dan intensitas pencahayaan [7]. Kombinasi HSV dan KNN telah diterapkan pada klasifikasi tingkat kematangan buah tomat dan mampu mencapai akurasi di atas 90%, meskipun kinerjanya masih dipengaruhi oleh kualitas citra dan variasi pencahayaan [8]. Temuan ini menegaskan bahwa kondisi akuisisi citra memiliki peran penting terhadap performa model klasifikasi berbasis warna.

Penelitian lain pada klasifikasi kematangan buah mangga berdasarkan citra HSV dan Fitur Statistik, menunjukkan bahwa algoritma KNN tetap mampu memberikan performa yang layak dengan akurasi sekitar 97,85% sedikit lebih baik dari algoritma SVM 97,50% [9]. Penelitian selanjutnya dengan judul Deteksi Kematangan Buah Jambu Kristal Berdasarkan Fitur Warna Menggunakan Metode Transformasi Ruang Warna Hsv (Hue Saturation Value) Dan *K-Nearest Neighbor* juga mendapatkan hasil akurasi yang mumpuni [10], selain itu penelitian dengan judul Penggunaan *K-Nearest Neighbor* (KNN) untuk Mengklasifikasi Citra Belimbing Berdasarkan Fitur Warna mendapatkan akurasi 93,33% [11], dari beberapa penelitian ini mengindikasikan bahwa kombinasi KNN dengan HSV memang memberikan performa yang mumpuni. Studi pada domain biologis lainnya mendapati akurasi KNN juga di pengaruhi oleh jumlah data latih karena mempengaruhi euclidian [12].

Di sisi lain, sebagian besar penelitian terdahulu lebih berfokus pada peningkatan akurasi melalui optimalisasi algoritma, pra-pemrosesan citra, atau teknik penyeimbangan data seperti *Synthetic Minority Over-sampling Technique* (SMOTE). Meskipun pendekatan tersebut terbukti efektif, dominasi rekayasa data dan model menyebabkan karakteristik alami dataset khususnya kondisi akuisisi citra relatif kurang dikaji sebagai variabel utama analisis.

Secara khusus, kajian yang membandingkan kinerja metode yang sama pada dataset dengan kondisi akuisisi citra yang berbeda, yaitu terkondisi (*controlled*) dan tidak terkondisi (*uncontrolled*), masih sangat terbatas. Mayoritas penelitian menggunakan citra dengan latar belakang dan pencahayaan yang seragam, sehingga performa model pada kondisi lapangan yang kompleks belum teruji secara komprehensif.

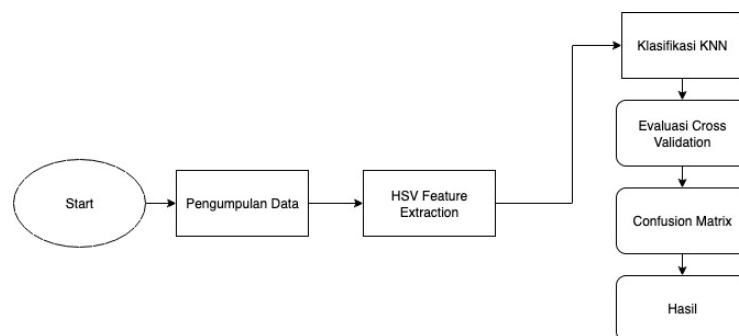
Berdasarkan celah penelitian tersebut, penelitian ini memiliki signifikansi dalam menganalisis secara komparatif kinerja klasifikasi citra butterfly fish pada dataset terkondisi dan tidak terkondisi menggunakan fitur HSV dan algoritma KNN tanpa penerapan teknik penyeimbangan data. Dengan mengunci metode ekstraksi fitur dan algoritma klasifikasi, penelitian ini menempatkan kondisi akuisisi citra sebagai variabel utama, sehingga memungkinkan evaluasi yang lebih objektif terhadap pengaruh kualitas, kuantitas, dan keragaman data terhadap performa model dalam konteks aplikasi nyata.

B. Metode Penelitian

Metode Penelitian eksperimen merupakan metode penelitian yang bertujuan untuk menguji pengaruh suatu perlakuan terhadap hasil yang ditimbulkan, sehingga hubungan sebab-akibat antar variabel dapat dianalisis secara sistematis. Dalam pendekatan ini, peneliti secara sengaja mengendalikan kondisi tertentu untuk mengamati perubahan yang terjadi akibat perlakuan yang diberikan. Keunggulan metode eksperimen terletak pada kemampuannya dalam menguji hubungan kausal secara langsung melalui pengendalian variabel penelitian [13].

Dalam konteks penelitian ini, pendekatan eksperimen digunakan untuk menganalisis pengaruh kondisi akuisisi citra terhadap kinerja klasifikasi. Pengujian dilakukan dengan membandingkan performa model pada dua kondisi dataset yang berbeda, yaitu dataset terkondisi dan tidak terkondisi, dengan parameter pengujian yang sama. Selain itu, pendekatan kuantitatif diterapkan melalui pengukuran kinerja model menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score. Dengan demikian, penelitian ini tidak hanya berfokus pada hasil klasifikasi, tetapi juga pada analisis hubungan antara karakteristik dataset khususnya kondisi akuisisi, jumlah data, dan keragaman visual terhadap performa algoritma *K-Nearest Neighbor* (KNN).

Penelitian ini dirancang secara sistematis untuk memastikan proses klasifikasi citra butterfly fish berjalan optimal dan menghasilkan evaluasi yang akurat. Tahapan penelitian dimulai dari pengumpulan data yang diperoleh melalui proses akuisisi citra, baik pada kondisi terkondisi maupun tidak terkondisi. Data yang terkumpul selanjutnya dilakukan ekstraksi fitur HSV guna memperoleh representasi warna yang menjadi karakteristik utama setiap spesies. Data hasil ekstraksi fitur kemudian diproses melalui tahap klasifikasi menggunakan algoritma KNN, yang selanjutnya dievaluasi menggunakan teknik cross-validation untuk mengukur performa model secara konsisten. Akurasi dan metrik evaluasi lainnya dianalisis melalui confusion matrix, sehingga diperoleh hasil akhir yang dapat digunakan untuk menilai efektivitas pendekatan yang diusulkan. Alur tahapan penelitian secara keseluruhan dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

1. Pengumpulan Data





Dataset yang digunakan dalam penelitian ini terdiri dari citra beberapa spesies butterfly fish yang diperoleh melalui proses akuisisi langsung oleh peneliti. Dataset dibagi menjadi dua kategori utama, yaitu dataset terkondisi dan dataset tidak terkondisi. Dataset terkondisi diambil dalam lingkungan mini studio dengan latar belakang kain putih untuk meminimalkan gangguan visual, menggunakan kamera DSLR Canon dengan resolusi citra sebesar 5184×3456 piksel. Dataset ini terdiri dari 9 kelas dengan total 191 citra, di mana jumlah citra pada kelas mayoritas mencapai 44 citra per kelas, sedangkan kelas minoritas memiliki hingga 9 citra per kelas.

Sementara itu, dataset tidak terkondisi diperoleh dalam kondisi alami, yaitu ikan difoto saat berenang bebas di dalam akuarium menggunakan kamera smartphone iPhone 11 dengan resolusi citra sebesar 4032×2268 piksel. Dataset ini terdiri dari 14 kelas dengan total 1555 citra, di mana kelas mayoritas memiliki hingga 307 citra per kelas dan kelas minoritas sebanyak 34 citra per kelas. Pemilihan perangkat smartphone dalam akuisisi data tidak terkondisi didasarkan pada skenario implementasi nyata (*real-world deployment*), di mana mayoritas pengguna akhir cenderung menggunakan perangkat mobile, mengingat kamera smartphone lebih unggul dari aspek keterjangkauan dan kemudahan operasional dibandingkan kamera profesional [13]. Dengan demikian, penggunaan smartphone iPhone 11 bertujuan untuk merepresentasikan kondisi akuisisi citra berbasis edge device, sehingga

model yang dikembangkan memiliki potensi generalisasi yang lebih baik serta kompatibilitas yang tinggi terhadap penggunaan di lingkungan operasional sehari-hari tanpa ketergantungan pada perangkat khusus.

Seluruh citra dalam penelitian ini digunakan dalam resolusi asli tanpa proses resizing maupun preprocessing awal, dengan tujuan mempertahankan informasi visual secara utuh, khususnya karakteristik warna berbasis HSV yang menjadi fokus utama penelitian. Meskipun terdapat perbedaan resolusi dan perangkat akuisisi, hal tersebut tidak memberikan pengaruh signifikan terhadap proses analisis karena fitur yang diekstraksi bersifat berbasis distribusi warna, bukan struktur spasial atau geometri objek. Pendekatan ini sekaligus memperkuat robustness model terhadap variasi kondisi akuisisi citra yang umum terjadi pada skenario penggunaan di dunia nyata. Berikut merupakan beberapa contoh sampel citra yang digunakan dalam dataset penelitian ini, yang merepresentasikan variasi visual antar kelas serta perbedaan kondisi akuisisi, baik pada dataset terkondisi (mini studio) maupun tidak terkondisi (*real-world environment*) menggunakan perangkat edge device (smartphone).

TABEL 1
 SAMPEL DATASET

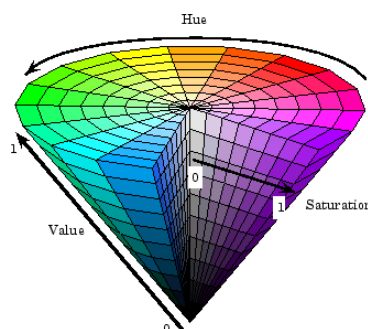
No	Spesies	Ciri-ciri	Citra
1	Kepe Kepe Coklat Strip	kuning kecoklatan, garis hitam pada mata, pola garis halus.	
2	Kepe Kepe Tiker	Abu-abu kekuningan, garis vertikal/diagonal kontras.	
3	Kepe Kepe Nanas	Kuning cerah, pola sisik seperti kulit nanas	
4	Kepe Kepe Tompel	Warnanya dominan kuning cerah. Terdapat tompel hitam besar di bagian tengah tubuh.	

Perlu diperhatikan bahwa terdapat perbedaan jumlah kelas dan jumlah citra antara dataset terkondisi dan dataset tidak terkondisi. Oleh karena itu, penelitian ini tidak dimaksudkan untuk melakukan perbandingan performa secara absolut, melainkan untuk mengamati kecenderungan pengaruh kondisi akuisisi citra terhadap kinerja model dengan tetap mempertimbangkan karakteristik alami masing-masing dataset. Selain itu, penelitian ini tidak menggunakan tahapan preprocessing tambahan seperti resizing, cropping, maupun filtering, dengan tujuan mempertahankan karakteristik asli citra sebagai representasi kondisi akuisisi yang sebenarnya.

2. Ekstraksi HSV (*Hue, Saturasi, Value*)

Model warna HSV (*Hue, Saturation, Value*) digunakan untuk merepresentasikan karakteristik warna pada citra. Fitur HSV direpresentasikan dalam bentuk vektor tiga dimensi (H, S, V) untuk setiap citra yang kemudian digunakan sebagai input pada proses klasifikasi. Hue menggambarkan jenis warna utama seperti merah, kuning, hijau, hingga biru, dengan rentang nilai antara 0 hingga 1. Saturation menunjukkan tingkat kejenuhan warna, di mana nilai mendekati 1 merepresentasikan warna yang lebih murni, sedangkan nilai mendekati 0 cenderung ke arah abu-abu. Value merepresentasikan tingkat kecerahan warna, dengan nilai yang lebih tinggi menunjukkan warna yang lebih terang [7].

Dalam penelitian ini, setiap citra awalnya berada dalam ruang warna RGB, kemudian dikonversi ke ruang warna HSV untuk memperoleh representasi warna yang lebih sesuai dengan persepsi visual manusia. Setelah proses konversi, nilai Hue, Saturation, dan Value dari setiap piksel dihitung, kemudian dirata-ratakan untuk menghasilkan tiga fitur utama yang merepresentasikan setiap citra.



Gambar 2. Hue Saturation Value

Pendekatan ini dipilih karena fitur warna HSV mampu menangkap perbedaan visual antar spesies ikan secara sederhana namun efektif, serta menghasilkan representasi fitur yang ringkas untuk proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Nilai piksel dalam ruang HSV selanjutnya digunakan sebagai input dalam proses klasifikasi [14]

3. Klasifikasi *K-Nearest Neighbor*

K-Nearest Neighbor (K-NN) merupakan metode klasifikasi yang mengelompokkan suatu objek berdasarkan kedekatannya terhadap data pelatihan yang memiliki karakteristik paling mirip. Proses klasifikasi dilakukan dengan menentukan sejumlah K tetangga terdekat, kemudian kelas mayoritas dari tetangga tersebut digunakan sebagai hasil prediksi. Nilai K berperan penting dalam kinerja model, di mana pemilihan K yang lebih besar cenderung mengurangi sensitivitas terhadap outlier, namun dapat menurunkan ketajaman dalam membedakan karakteristik spesifik, termasuk representasi warna yang menjadi ciri utama setiap spesies [15]. Data hasil ekstraksi fitur selanjutnya diproses menggunakan perhitungan jarak sebagai dasar penentuan kedekatan antar data. Dalam algoritma K-NN, pengukuran jarak dilakukan menggunakan Euclidean Distance, yang mengkuantifikasi jarak antar titik dalam ruang fitur. Semakin kecil nilai jarak yang dihasilkan, semakin tinggi tingkat kemiripan antara data uji dan data latih [16]. Oleh karena itu, kualitas klasifikasi sangat ditentukan oleh kemampuan model dalam mengidentifikasi data dengan jarak minimum terhadap data uji. Perhitungan jarak tersebut dilakukan menggunakan rumus Euclidean Distance sebagaimana dirumuskan pada persamaan berikut [17]

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_{training}^i - y_{testing}^i)^2} \quad (1)$$

Keterangan:

$d(x,y)$ = Jarak antara data training dan data testing

$x_{training}^i$ = Data training

$y_{testing}^i$ = Data testing

i = Variabel data

n = Dimensi data

Algoritma *K-Nearest Neighbor* (K-NN) banyak diadopsi dalam klasifikasi citra karena kemampuannya dalam memanfaatkan struktur lokal data tanpa memerlukan proses pelatihan yang kompleks. Metode ini mengevaluasi kemiripan berdasarkan jarak dalam ruang fitur, sehingga secara langsung sensitif terhadap distribusi fitur warna. Oleh karena itu, K-NN menjadi pendekatan yang relevan ketika informasi warna berperan dominan dalam membedakan antar kelas [18]. Nilai K yang digunakan dalam penelitian ini adalah 3, 5, 7, dan 9 dipilih untuk menguji sensitivitas dan konsistensi performa model terhadap perubahan jumlah tetangga. Penggunaan nilai K ganjil secara eksplisit ditujukan untuk mengeliminasi ambiguitas keputusan akibat kemungkinan voting yang seimbang, sehingga hasil klasifikasi selalu ditentukan oleh mayoritas yang jelas. Variasi ini juga digunakan untuk mengamati trade-off antara ketahanan terhadap noise dan kemampuan model dalam mempertahankan representasi lokal data. [19]

4. Evaluasi

Confusion matrix digunakan untuk membandingkan hasil prediksi model dengan kelas aktual pada proses klasifikasi citra butterfly fish. Melalui matriks ini, dapat diamati distribusi prediksi benar dan salah pada setiap kelas, sehingga memberikan gambaran yang lebih rinci terhadap kinerja model *K-Nearest Neighbor* (KNN) yang

digunakan. Berdasarkan confusion matrix, selanjutnya dihitung metrik evaluasi berupa akurasi, precision, recall, dan F1-score untuk mengukur performa model secara kuantitatif. [20]

Untuk memastikan hasil evaluasi yang lebih stabil dan tidak bergantung pada pembagian data tertentu, penelitian ini menggunakan skema k-fold cross-validation dengan nilai k = 5 dan k = 10. Pada skema ini, dataset dibagi menjadi k bagian, di mana satu bagian digunakan sebagai data uji dan sisanya sebagai data latih. Proses ini diulang sebanyak k kali hingga seluruh data berperan sebagai data uji secara bergantian, dan nilai evaluasi akhir diperoleh dari rata-rata seluruh iterasi. Pendekatan ini digunakan untuk mengurangi bias akibat pembagian data tunggal dan meningkatkan reliabilitas hasil evaluasi. [21]

III. HASIL DAN PEMBAHASAN

A. Confusion Matrix

Pada tahap awal evaluasi, performa model KNN dianalisis menggunakan confusion matrix sebagai dasar untuk menghitung metrik evaluasi seperti akurasi, precision, recall, dan F1-score. Confusion matrix memberikan representasi distribusi hasil klasifikasi dengan membandingkan antara kelas aktual dan kelas hasil prediksi. [22]

	true Coklat strip	true Tiker	true Melanotus	true Orega	true Gabus	true Panah	true Monyong	true Monyong asli	true Nanas
pred. Coklat strip	7	3	1	1	2	0	0	0	0
pred. Tiker	0	3	0	0	0	2	0	1	0
pred. Melanotus	0	1	4	1	1	2	1	0	0
pred. Orega	0	0	1	3	0	0	0	0	0
pred. Gabus	2	2	1	0	6	1	0	0	0
pred. Panah	0	0	1	2	0	4	0	0	0
pred. Monyong	0	0	1	2	0	0	6	0	0
pred. Monyong ...	0	0	0	0	0	0	0	7	2
pred. Nanas	0	0	0	0	0	0	2	1	7

Gambar 3. Contoh Hasil Confusion Matrix

Gambar 3 menunjukkan contoh confusion matrix yang dihasilkan dari proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan nilai K = 5 pada skema 5-fold cross-validation. Confusion matrix ini ditampilkan sebagai representasi hasil klasifikasi, mengingat jumlah kombinasi pengujian yang cukup banyak sehingga tidak seluruhnya ditampilkan dalam penelitian ini. Pada confusion matrix, nilai pada diagonal utama menunjukkan jumlah prediksi yang benar, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi antar kelas. Informasi ini menjadi dasar dalam menghitung metrik evaluasi yang digunakan untuk menilai kinerja model secara kuantitatif. Berdasarkan confusion matrix tersebut, selanjutnya dilakukan perhitungan nilai akurasi, precision, recall, dan F1-score untuk memperoleh gambaran performa model secara keseluruhan.

B. Hasil Uji Dataset Pada Model KNN (*K-Nearest Neighbor*)

Pengujian data terkondisi dan data tidak terkondisi dilakukan menggunakan KNN dengan nilai k = 3, 5, 7, dan 9, serta skema validasi k-fold 5 dan 10.

TABEL 2
HASIL UJI DATASET MENGGUNAKAN K-FOLD 5

K Fold	Nilai K	Metode Data	Akurasi	Precision	Recall	F1-score
5	3	Dataset Terkondisi	64.21%	69.64%	57.89%	56.91%
5	5	Dataset Terkondisi	66.31%	76.63%	59.64%	59.53%
5	7	Dataset Terkondisi	63.68%	78.22%	66.66%	67.50%
5	9	Dataset Terkondisi	63.68%	77.43%	63.15%	62.92%
5	3	Dataset Tidak Terkondisi	74.52%	73.03%	73.23%	72.40%
5	5	Dataset Tidak Terkondisi	75.03%	75.01%	74.30%	73.31%
5	7	Dataset Tidak Terkondisi	75.29%	75.72%	74.94%	73.97%
5	9	Dataset Tidak Terkondisi	74.77%	75.56%	74.94%	73.99%

TABEL 3
HASIL UJI DATASET MENGGUNAKAN K-FOLD 10

K Fold	Nilai K	Metode Data	Akurasi	Precision	Recall	F1-score
10	3	Dataset Terkondisi	67.89%	69.64%	57.89%	56.91%
10	5	Dataset Terkondisi	66.31%	76.63%	59.64%	59.53%
10	7	Dataset Terkondisi	66.84%	78.22%	66.66%	67.50%
10	9	Dataset Terkondisi	64.21%	77.43%	63.15%	62.92%
10	3	Dataset Tidak Terkondisi	75.03%	73.03%	73.23%	72.40%
10	5	Dataset Tidak Terkondisi	75.61%	75.00%	74.30%	73.31%
10	7	Dataset Tidak Terkondisi	76.00%	75.72%	74.94%	73.97%
10	9	Dataset Tidak Terkondisi	75.61%	75.56%	74.94%	73.99%

Dari Tabel I dan Tabel II dapat dilihat bahwa hasil eksperimen menunjukkan dataset tidak terkondisi cenderung menghasilkan nilai akurasi, recall, dan F1-score yang lebih tinggi dibandingkan dataset terkondisi, baik pada skema k-fold 5 maupun k-fold 10 dan pada seluruh variasi nilai k. Di sisi lain, nilai precision pada dataset tidak terkondisi cenderung sedikit lebih rendah dibandingkan dataset terkondisi. Pola ini menunjukkan adanya trade-off antara kemampuan model dalam menjangkau seluruh kelas (recall) dan ketepatan prediksi pada kelas tertentu (precision).

Peningkatan performa pada dataset tidak terkondisi tidak hanya dipengaruhi oleh variasi visual yang lebih tinggi, tetapi juga oleh jumlah data yang jauh lebih besar. Dengan 1.555 citra, distribusi sampel pada dataset tidak terkondisi menjadi lebih padat dan representatif di ruang fitur, sehingga algoritma KNN mampu menemukan tetangga terdekat yang lebih stabil secara statistik. Kepadatan ini berkontribusi terhadap peningkatan akurasi dan F1-score, karena keputusan klasifikasi tidak lagi bergantung pada sejumlah kecil sampel yang rentan terhadap noise atau bias lokal.

Sebaliknya, keterbatasan jumlah data pada dataset terkondisi menyebabkan kepadatan ruang fitur menjadi rendah, terutama pada kelas minoritas. Meskipun kondisi akuisisi yang seragam menghasilkan citra dengan karakteristik visual yang konsisten dan meningkatkan precision, model cenderung mengalami overfitting lokal, sehingga kurang mampu mengenali variasi baru. Hal ini menjelaskan nilai recall dan F1-score yang lebih rendah pada dataset terkondisi.

Selain itu, dataset tidak terkondisi memiliki variasi visual yang lebih luas, mencakup perubahan pose, orientasi tubuh, pencahayaan, dan latar belakang. Kombinasi antara jumlah data yang besar dan keragaman visual memungkinkan model KNN membentuk distribusi fitur yang lebih komprehensif dan mendekati kondisi nyata di lapangan. Dampaknya, kemampuan generalisasi meningkat, yang tercermin dari kenaikan nilai recall. Namun, kondisi ini juga meningkatkan tumpang tindih antar kelas di ruang fitur, sehingga probabilitas false positive meningkat, yang menjelaskan penurunan ringan pada precision.

Analisis terhadap variasi nilai k menunjukkan bahwa peningkatan jumlah tetangga dari k=3 ke k=9 cenderung menurunkan performa, khususnya pada akurasi dan F1-score. Fenomena ini disebabkan oleh semakin banyaknya tetangga dari kelas berbeda yang dipertimbangkan dalam proses klasifikasi, terutama pada spesies butterfly fish yang memiliki kemiripan pola warna dan tekstur yang tinggi. Pola ini konsisten pada kedua jenis dataset, namun dampaknya lebih tereduksi pada dataset tidak terkondisi karena kepadatan data yang lebih tinggi.

Dari sisi skema validasi, penggunaan k-fold 10 menghasilkan performa yang lebih stabil dibandingkan k-fold 5, terutama pada dataset tidak terkondisi. Stabilitas ini menunjukkan bahwa jumlah data yang lebih besar mampu mendukung pembagian data latih dan uji yang lebih representatif. Sebaliknya, pada dataset terkondisi, fluktuasi performa antar lipatan masih terlihat, mengindikasikan bahwa jumlah data yang terbatas belum sepenuhnya memadai untuk menghasilkan estimasi performa yang robust.

Secara keseluruhan, hasil ini menunjukkan bahwa jumlah dan keragaman data memberikan kontribusi yang signifikan terhadap kinerja KNN, di samping pengaruh kondisi akuisisi citra. Namun demikian, temuan ini perlu diinterpretasikan secara hati-hati karena perbedaan jumlah data dan jumlah kelas antara kedua dataset juga berpotensi memengaruhi hasil perbandingan.

IV. SIMPULAN

Penelitian ini mengevaluasi kinerja metode *K-Nearest Neighbor* (KNN) dalam klasifikasi citra beberapa spesies butterfly fish menggunakan dataset terkondisi dan tidak terkondisi dengan variasi nilai k serta skema validasi k -fold 5 dan 10. Hasil eksperimen menunjukkan bahwa dataset tidak terkondisi, yang memiliki jumlah data lebih besar dan variasi visual lebih tinggi, cenderung menghasilkan nilai akurasi, recall, dan F1-score yang lebih baik dibandingkan dataset terkondisi. Meskipun terdapat sedikit penurunan pada nilai precision, secara umum performa model pada dataset tidak terkondisi menunjukkan kemampuan generalisasi yang lebih baik.

Namun demikian, hasil tersebut perlu diinterpretasikan secara hati-hati, mengingat adanya perbedaan jumlah data dan jumlah kelas antara dataset terkondisi dan tidak terkondisi yang dapat memengaruhi perbandingan performa secara langsung. Oleh karena itu, peningkatan performa yang diperoleh tidak semata-mata disebabkan oleh kondisi akuisisi citra, tetapi juga dipengaruhi oleh karakteristik dataset, khususnya jumlah dan keragaman data. Penelitian ini juga memiliki keterbatasan pada ketidakseimbangan jumlah data antar kelas, terutama pada kelas minoritas, serta penggunaan metode klasifikasi berbasis jarak yang belum mampu melakukan pembelajaran fitur secara adaptif.

Sebagai pengembangan ke depan, penelitian selanjutnya dapat mempertimbangkan penerapan teknik penyeimbangan data seperti SMOTE, serta melakukan perbandingan dengan metode berbasis deep learning yang memiliki kemampuan ekstraksi fitur otomatis dan lebih robust terhadap variasi visual. Selain itu, eksplorasi fitur yang lebih kompleks serta pengujian pada skenario data yang lebih sebanding diharapkan dapat memberikan analisis yang lebih komprehensif dan meningkatkan keandalan sistem klasifikasi ikan untuk aplikasi dunia nyata.

DAFTAR PUSTAKA

- [1] B. Arbanto, A. Y. Retawimbi, A. Faricha And G. P. S. Rifani, "Ikan Karang Famili Chaetodontidae Di Kepulauan Ayau, Kabupaten Raja Ampat, Papua Barat," *Jurnal Enggano*, Vol. 5, No. 2, Pp. 195-204, 2020.
- [2] J. S. N. Silalahi, Anisah, R. Fitra And C. Rozikin, "Penggunaan Ciri Warna Hsv Pada Bola Mata Ikan Untuk Identifikasi Tingkat Kesegaran Ikan Menggunakan Algoritma Knn," *Jurnal Teknologi Informasi*, Vol. 02, No. 02, Pp. 85-90, 2023.
- [3] A. Paliling, M. Muchtar And F. Fardian, "Sistem Cerdas Deteksi Kematangan Buah Naga Berbasis Hsv-Knn," *Jurnal Sistem Informasi Dan Teknologi Informasi*, Vol. 14, No. 1, Pp. 46-55, 2025.
- [4] S. Sanjaya, M. L. Pura, S. K. Gusti, F. Yanto And F. Syafria, "*K-Nearest Neighbor* Untuk Klasifikasi Tingkat Kematangan Tomat Berdasarkan Hue, Saturation, And Value Colors," *Indonesian Journal Of Artificial Intelligence And Data Mining (Ijaidm)*, Vol. 2, No. 2, Pp. 101-106, 2019.
- [5] F. Shidiq, E. W. Hidayat And N. I. Kurniati, "Penerapan Metode *K-Nearest Neighbor* (Knn) Untuk Menentukan Ikan Cupang Menggunakan Deteksi Tepi Canny Dan Invariant Moment," *Jurnal Teknik Informatika (Jutif)*, Vol. 3, No. 1, Pp. 11-20, 2022.
- [6] I. P. A. Putra, I. M. A. Wirawan And I. G. A. Gunadi, "Peningkatan Akurasi Klasifikasi Citra Butterfly Fish Menggunakan Ekstraksi Fitur Hsv Dan Penyeimbangan Data Berbasis Smote," *Jurnal Inovtek Polbeng -Seri Informatika*, Vol. 10, No. 3, Pp. 1402-1411, 2025.
- [7] C. N. Mayasari, Pujiono And M. Soeleman, "Deteksi Pornografi Pada Karakter Animasi 2d Dengan Knn (*K-Nearest Neighbors*) Menggunakan Fitur Hsv," *Jurnal Penelitian Dan Pengabdian Masyarakat*, Vol. 2, No. 8, Pp. 1236-1250, 2022.
- [8] L. Ningsih And P. Cholidhazia, "Klasifikasi Tingkat Kematangan Tomat Berdasarkan Warna Rgb Dan Hsv Menggunakan Algoritma Knn," *Journal Of Artificial Intelligence And Digital Business*, Vol. 1, No. 1, Pp. 25-30, 2022.
- [9] M. Muchtar And R. A. Muchtar, "Perbandingan Metode Knn Dan Svm Dalam Klasifikasi Kematangan Buah Mangga Berdasarkan Citra Hsv Dan Fitur Statistik," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 12, No. 2, Pp. 876-884, 2024.
- [10] A. Wibowo, D. M. C. Hermanto, K. I. Lestari And H. Wijoyo, "Deteksi Kematangan Buah Jambu Kristal Berdasarkan Fitur Warna Menggunakan Metode Transformasi Ruang Warna Hsv (Hue Saturation Value) Dan *K-Nearest Neighbor*," *Jurnal Of Informatics And Computer Science Engineering*, Pp. 76-88, 2021.
- [11] D. I. Muhammad, Ermatita And N. Falih, "Penggunaan *K-Nearest Neighbor* (Knn) Untuk Mengklasifikasi Citra Belimbing Berdasarkan Fitur Warna," *Jurnal Informatik*, Vol. 17, No. 1, Pp. 9-16, 2021.
- [12] M. R. Siregar, N. Hidayat And R. K. Dewi, "Implementasi Metode Modified *K-Nearest Neighbor* (Mk-Nn) Untuk Diagnosis Penyakit Tanaman Kentang," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, Vol. 5, No. 8, Pp. 3217-3221, 2021.
- [13] J. S. Saliama And T. Wibowo, "Studi Komparasi Teknik Antara Dslr Dan Smartphone Photography," *Jurnal Universitas Internasional Batam*, Vol. 1, No. 1, Pp. 439-451, 2020.
- [14] A. Kadir And A. Susanto, *Teori Dan Aplikasi Pengolahan Citra*, Yogyakarta: Andi, 2013.
- [15] A. Yudhana, Sunardi And D. A. J. S. Hartanta, "Algoritma K-Nn Dengan Euclidean Distance Untuk Prediksi Hasil Penggajian Kayu Sengon," *Algoritma K-Nn Dengan Euclidean Distance Untuk Prediksi Hasil Penggajian Kayu Sengon*, Vol. 22, No. 04, Pp. 123-129, 2020.
- [16] C. S. A. Nara, I. M. D. Ardiada And P. W. Rahayu, "Rancang Bangun Sistem Rekomendasi Tempat Kuliner Di Desa Cunggu Menggunakan Algoritma *K-Nearest Neighbor*," *Jurnal Mahasiswa Teknik Informatika*, Vol. 10, No. 2, Pp. 2408-2415, 2026.
- [17] A. F. Pulungan And M. Zarlis, "Analysis Of Braycurtis, Canberra And Euclidean Distance In Knn Algorithm," *Journal Publications & Informatics Engineering Research*, Vol. 4, No. 1, Pp. 74-77, 2019.

- [18] A. Sugesti, M. A. A. Mukid And Tarno, "Perbandingan Kinerja Mutual *K-Nearest Neighbor*(Mknn) Dan *K-Nearest Neighbor*(Knn) Dalam Analisis Klasifikasi Kelayakan Kredit," *Jurnal Gaussian*, Vol. 8, No. 3, Pp. 366-376, 2019.
- [19] M. R. Azhar, A. Iskandar And A. Rifqi, "Klasifikasi Algoritma *K- Nearest Neighbors* (Knn) Untuk Segmentasi Pasar Mobil Bekas Berdasarkan Merek Dan Harga," (*Jurnal Mahasiswa Teknik Informatika*, Vol. 9, No. 3, Pp. 3881-3888, 2025.
- [20] A. F. Sadeli And I. I. Lawanda, "Recall, Precision, Dan F-Measure Untuk Mengevaluasi Sistem Temu Kembali Informasi Pada Sistem Manajemen Dokumen Elektronik (Edms)" [Recall, Precision, And F-Measure For Evaluating Information Retrieval System In Electronic Document Management Systems (Edm)," *Jurnal Ilmu Perpustakaan, Informasi, & Kearsipan*, Vol. 11, No. 2, Pp. 231-241, 2023.
- [21] H. Hafid, "Penerapan *K-Fold Cross Validation* Untuk Menganalisis Kinerja Algoritma *K-Nearest Neighbor* Pada Data Kasus Covid-19 Di Indonesia," *Journal Of Mathematics, Computations, And Statistics*, Vol. 6, No. 2, Pp. 161-168, 2023.
- [22] L. Hakim, A. Sobri, L. Sunardi And D. Nurdiansyah, "Prediksi Penyakit Jantung Berbasis Mesin Learning Dengan Menggunakan Metode *K-Nn*," *Jurnal Digital Teknologi Informasi*, Vol. 07, No. 02, Pp. 14-20, 2024.
- [23] M. F. Arib, M. S. Rahayu, R. A. Sidorj And M. W. Afgani, "Experimental Research Dalam Penelitian Pendidikan," *Innovative: Journal Of Social Science Research*, Vol. 4, No. 1, Pp. 5497-5511, 2024.