

IMPROVING BUTTERFLY FISH IMAGE CLASSIFICATION ACCURACY USING HSV FEATURE EXTRACTION AND SMOTE-BASED DATA BALANCING

PENINGKATAN AKURASI KLASIFIKASI CITRA BUTTERFLY FISH MENGGUNAKAN EKSTRAKSI FITUR HSV DAN PENYEIMBANGAN DATA BERBASIS SMOTE

I Putu Arya Putra¹, I Made Agus Wirawan², I Gede Aris Gunadi³

¹Ilmu Komputer, Program Pascasarjana,

²Jurusan Teknik Informatika, Fakultas Teknik dan Vokasi,

³Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Pendidikan Ganesha, Singaraja, Indonesia

arya.putra.2@student.undiksha.ac.id¹, imade.aguswirawan@undiksha.ac.id²,
igedearisgunadi@undiksha.ac.id³

Abstract – Class imbalance in image data can reduce the accuracy of classification models, especially when the minority class data is much smaller than the majority class. This research focuses on enhancing the classification accuracy of butterflyfish images through the application of the Synthetic Minority Over-sampling Technique (SMOTE) for data balancing, combined with the K-Nearest Neighbor (KNN) algorithm utilizing HSV-based feature extraction. The datasets were collected in two conditions, namely conditioned (controlled background and lighting) and unconditioned (varied background and natural lighting). The research stages include pre-processing, HSV feature extraction, data balancing with SMOTE, and classification using KNN with various k values (3, 5, 7, 9) and cross-validation (k -fold 5 and 10). The experimental results show that SMOTE consistently improves accuracy on both types of datasets, with the best performance at $k = 3$ and k -fold = 10, namely 85.32% (conditioned) and 87.59% (unconditioned). This improvement occurs because a more balanced data distribution allows the model to optimally recognise features between classes. This study proves that the integration of SMOTE and KNN is effective in overcoming class imbalance in image classification, with potential applications in the fields of digital image technology, ecosystem management, and species identification.

Keywords – Butterfly Fish, Image Classification, HSV, SMOTE, Data Imbalance.

Abstrak - Ketidakseimbangan kelas pada data citra dapat menurunkan akurasi model klasifikasi, khususnya ketika jumlah data kelas minoritas jauh lebih sedikit daripada kelas mayoritas. Penelitian ini berfokus pada peningkatan akurasi klasifikasi citra butterfly fish dengan menerapkan metode Synthetic Minority Over-sampling Technique (SMOTE) untuk penyeimbangan data, serta algoritma K-Nearest Neighbor (KNN) yang dioptimalkan melalui ekstraksi fitur pada ruang warna HSV. Dataset dikumpulkan dalam dua kondisi, yaitu terkondisi (latar dan pencahayaan terkontrol) dan tidak terkondisi (variasi latar dan pencahayaan alami). Tahapan penelitian meliputi pra-pemrosesan, ekstraksi fitur HSV, penyeimbangan data dengan SMOTE, dan klasifikasi menggunakan KNN dengan variasi nilai k (3, 5, 7, 9) serta validasi silang (k -fold 5 dan 10). Hasil eksperimen menunjukkan SMOTE secara konsisten meningkatkan akurasi pada kedua jenis dataset, dengan performa terbaik pada $k=3$ dan k -fold=10, yaitu 85,32% (terkondisi) dan 87,59% (tidak terkondisi). Peningkatan ini terjadi karena distribusi data yang lebih seimbang memungkinkan model mengenali fitur antar kelas secara optimal. Penelitian ini membuktikan bahwa integrasi SMOTE dan KNN efektif mengatasi class imbalance pada klasifikasi citra, dengan potensi penerapan di bidang teknologi citra digital, pengelolaan ekosistem, dan identifikasi spesies.

Kata Kunci – Butterfly Fish, Klasifikasi Citra, HSV, SMOTE, Data Imbalance.

I. PENDAHULUAN

Butterfly Fish merupakan salah satu ikan karang tropis dengan tampilan visual mencolok dan nilai ekonomi tinggi di industri ikan hias. Identifikasi spesies yang akurat sangat penting untuk mendukung konservasi ekosistem terumbu karang dan mengatur perdagangan ikan hias[1]. Metode identifikasi konvensional selama ini masih bergantung pada keahlian pakar, yang bersifat subjektif, memakan waktu lama, dan sulit diakses oleh masyarakat umum tanpa pengetahuan khusus. Perkembangan machine learning dan pengolahan citra digital telah membuka peluang identifikasi berbasis citra yang lebih cepat, konsisten, dan dapat diotomatisasi.

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 pentingnya kualitas gambar terhadap kinerja model [2]. Studi lain pada klasifikasi ikan air tawar dengan jumlah data terbatas (100 citra dari 10 jenis ikan) menunjukkan bahwa KNN tetap mampu memberikan hasil yang layak, meskipun akurasi hanya sekitar 70%, mengindikasikan bahwa algoritma sederhana ini masih relevan untuk dataset berskala kecil[3]. Di sisi lain, masalah ketidakseimbangan kelas (*class imbalance*) telah menjadi tantangan umum yang dapat mengurangi performa model klasifikasi. Untuk mengatasinya, beberapa penelitian memanfaatkan *Synthetic Minority Over-sampling Technique* (SMOTE) dengan hasil positif. Dalam domain citra medis, integrasi SMOTE dengan pendekatan transfer learning dilaporkan mampu meningkatkan akurasi sebesar 1,97%, recall sebesar 5,43%, dan specificity sebesar 0,72%, membuktikan efektivitas metode ini dalam meningkatkan kinerja model pada data yang tidak seimbang[4].

Meskipun berbagai penelitian tersebut membuktikan efektivitas HSV, KNN, dan SMOTE, integrasi ketiganya pada kasus klasifikasi Butterfly Fish di dua kondisi akuisisi—terkondisi (controlled) dan tidak terkondisi (uncontrolled)—masih jarang dilakukan. Sebagian penelitian terdahulu cenderung hanya menggunakan data seimbang atau citra dengan kondisi terkendali, sehingga kurang merepresentasikan situasi lapangan yang kompleks. Hal ini menjadi celah penelitian yang penting untuk diisi, mengingat kondisi pencahayaan, latar belakang, dan posisi ikan dapat memengaruhi hasil klasifikasi. Penelitian ini mengusulkan peningkatan performa klasifikasi citra spesies butterfly fish dengan menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE merupakan metode resampling yang banyak digunakan untuk menyeimbangkan distribusi data dengan cara menghasilkan data sintesis untuk kelas minoritas melalui interpolasi terhadap data sejenis di sekitarnya[5]. Penelitian ini difokuskan untuk menganalisis pengaruh penerapan SMOTE terhadap akurasi klasifikasi citra *butterfly fish* berbasis HSV dan membandingkan performa klasifikasi pada data terkondisi dan tidak terkondisi menggunakan model KNN dengan berbagai nilai k dan skema validasi.

II. SIGNIFIKANSI STUDI

A. Penelitian Terdahulu

Berbagai penelitian terdahulu telah mengeksplorasi penerapan metode *Synthetic Minority Over-sampling Technique (SMOTE)* dan *algoritma K-Nearest Neighbor (KNN)* pada beragam domain data. Terdapat penelitian memanfaatkan SMOTE pada data Car Evolution dari UCI Machine Learning Repository dan berhasil meningkatkan akurasi rata-rata KNN sebesar 9,97%, dengan kinerja terbaik pada nilai $k=3$ (93,11%)[6]. Hasil ini menegaskan bahwa penanganan ketidakseimbangan kelas mampu memperbaiki performa model, terutama pada skenario di mana proporsi kelas sangat timpang.

Pada domain industri, menerapkan SMOTE untuk mengatasi ketidakseimbangan data evaluasi periklanan, yang berhasil meningkatkan akurasi sebesar 0,73% dan menurunkan loss baik pada tahap pelatihan maupun validasi[7]. Penelitian pada pengukuran kecepatan angin menekankan pentingnya akurasi sistem sensor, meskipun tidak secara langsung menggunakan SMOTE, namun relevan dalam konteks peningkatan kualitas data[8]. Pada bidang pengolahan citra digital, menerapkan ekstraksi fitur geometri pada citra kendaraan dan mengklasifikasikannya menggunakan KNN dengan akurasi 89%[9]. Selanjutnya terdapat klasifikasi tingkat kematangan tomat menggunakan KNN, mencapai akurasi 91,25%, meskipun dipengaruhi oleh variasi intensitas cahaya[10]. Penelitian memanfaatkan fitur warna HSV untuk mendeteksi kesegaran ikan dengan akurasi uji 90%[11]. Selanjutnya penelitian mengklasifikasikan tanaman hias berbasis HSV-KNN dengan akurasi 92%, menekankan perlunya pra-pemrosesan untuk mengatasi perbedaan pencahayaan[12].

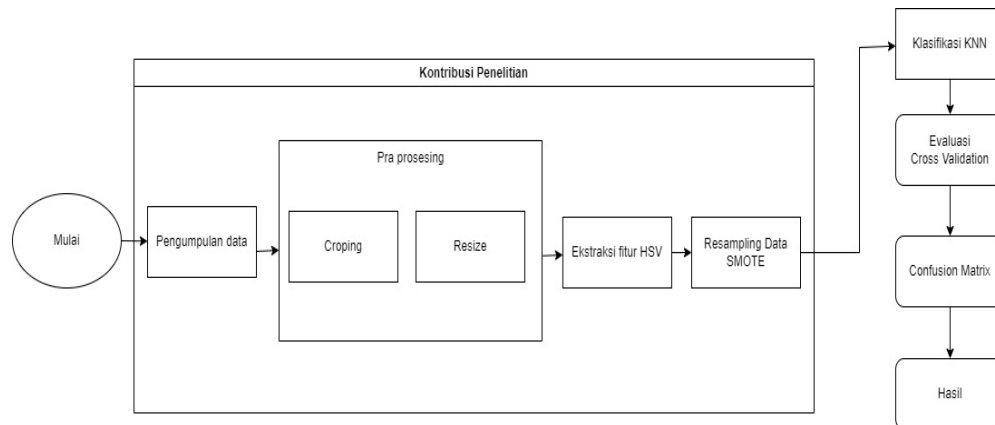
Rangkaian penelitian ini memperlihatkan bahwa penggabungan teknik pra-pemrosesan citra, ekstraksi fitur berbasis warna HSV, serta algoritma KNN yang didukung metode penyeimbangan data seperti SMOTE, memiliki relevansi luas. Pendekatan ini tidak hanya bermanfaat pada sektor spesifik seperti pertanian (penentuan kematangan buah) atau perikanan (identifikasi kesegaran ikan), tetapi juga dapat diadaptasi untuk pengelolaan ekosistem, pemantauan kualitas lingkungan, sistem inspeksi industri, dan aplikasi keamanan digital. Teknologi ini mendukung otomasi pengambilan keputusan berbasis citra, yang penting untuk efisiensi, konsistensi, dan kecepatan analisis di berbagai bidang. Meskipun metode HSV-KNN dan SMOTE telah terbukti efektif pada berbagai jenis data, sebagian besar studi terdahulu memiliki keterbatasan:

1. Banyak penelitian hanya fokus pada satu jenis kondisi akuisisi citra (misalnya latar seragam), sehingga performa pada kondisi lapangan yang tidak terkendali belum teruji secara optimal.
2. Analisis jarang membandingkan performa metode pada dataset terkondisi dan tidak terkondisi secara bersamaan, sehingga belum ada gambaran komprehensif terkait robustnes model.
3. Evaluasi bias yang mungkin muncul akibat SMOTE, seperti potensi *overfitting* atau *sampling bias* pada data sintesis, belum dibahas secara mendalam

B. Metode Penelitian

Metode 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 kemudian memasuki tahap pra-pemrosesan, yang meliputi *cropping* untuk memfokuskan area objek utama, serta *resize* untuk menyeragamkan dimensi citra. Selanjutnya, dilakukan ekstraksi fitur HSV guna memperoleh representasi warna yang menjadi karakteristik utama setiap spesies. Untuk mengatasi permasalahan ketidakseimbangan kelas, diterapkan metode resampling data menggunakan SMOTE yang

menghasilkan sampel sintetis pada kelas minoritas. Data hasil resampling 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. Dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

1. Pengumpulan Data

Dataset yang dijadikan bahan pengujian dalam studi ini merupakan citra beberapa spesies butterfly fish yang diperoleh langsung melalui proses akuisisi lapangan oleh peneliti. Data dikumpulkan dalam dua kondisi: (1) data terkondisi, yaitu citra ikan yang diambil menggunakan kamera Canon dalam mini-studio dengan alas kain putih sebagai latar belakang; serta (2) data tidak terkondisi, yaitu citra ikan yang difoto menggunakan kamera iPhone 11 saat ikan berenang bebas dalam akuarium. Pendekatan ini memungkinkan variasi visual yang representatif terhadap kondisi nyata di lapangan. Dataset ini terdiri dari 9 kelas untuk data terkondisi, dengan jumlah data pada kelas mayoritas sebanyak 44 dan kelas minoritas sebanyak 9 gambar, serta 14 kelas untuk data tidak terkondisi, dengan jumlah data kelas mayoritas sebanyak 307 dan kelas minoritas 34 gambar. Dapat dilihat pada Tabel 1.

TABEL I
CONTOH VISUAL BEBERAPA SPESIES BUTTERFLYFISH

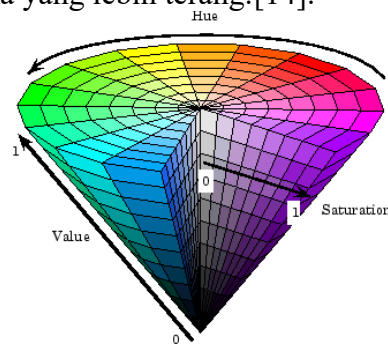
No	Spesies	Ciri Ciri	Citra
1	Kepe Coklat	Spesies ini memiliki morfologi tubuh yang relatif lebar, dilapisi sisik berstruktur estetik, dengan dominasi warna kuning kecoklatan serta garis hitam yang melintang pada area mata	
2	Kepe Tiker	Spesies ini memiliki warna dominan abu-abu dengan pola garis-garis sejajar pada tubuhnya, disertai aksen kekuningan pada sirip dan area tubuh bagian bawah.	
3	Kepe Nanas	Spesies Kepe-kepe Nanas memiliki karakteristik warna kuning cerah dan pola sisik yang menyerupai permukaan buah nanas, sehingga mudah dikenali secara visual.	
4	Kepe Monyong	Spesies ini dicirikan oleh bentuk mulut yang menonjol ke depan, dengan kombinasi warna putih pada area mulut serta pola garis berwarna oranye pada tubuhnya.	

2. *Preprocessing*

Setiap citra yang terkumpul diproses dalam tahap prapemrosesan, termasuk pemotongan (cropping), pengaturan dimensi (resize), dan pemberian label. Seluruh data distandarisasi ke ukuran 224×224 piksel untuk menjaga keseragaman[13].

3. *Ekstraksi HSV (Hue, saturasi, value)*

Pada model HSV, warna direpresentasikan melalui tiga parameter utama: Hue, Saturation, dan Value. Hue mendefinisikan kategori warna, seperti merah, ungu, atau kuning, serta menentukan perbedaan antar warna dan tingkat kecenderungan cahaya ke arah merah atau hijau. Komponen Hue memiliki rentang nilai antara 0 hingga 1, yang merepresentasikan perubahan warna secara berurutan dari merah, kuning, hijau, sian, biru, magenta, lalu kembali ke merah. Sementara itu, Saturation juga berada pada skala 0 hingga 1 dan menunjukkan tingkat kemurnian warna, dari warna yang sangat jenuh tanpa campuran abu-abu hingga warna pucat yang mendekati putih. Value, atau luminansi, juga memiliki rentang 0–1, dan menggambarkan intensitas kecerahan, di mana angka yang lebih tinggi menandakan warna yang lebih terang.[14].



Gambar 2. Hue Saturatin Value

Suatu model warna dapat dipandang sebagai sistem koordinat dengan subdivisi ruang di dalamnya, di mana setiap warna direpresentasikan sebagai sebuah titik. Ruang warna ini dirancang untuk menyediakan standar yang tegas dalam penentuan spesifikasi warna. Dalam proses dekomposisi citra berwarna RGB, nilai piksel memegang peranan krusial sebelum dilakukan konversi ke ruang warna HSV.[15].

4. *Resampling Data SMOTE (Synthetic Minority Oversampling Technique)*

SMOTE merupakan salah satu metode oversampling yang banyak digunakan untuk mengatasi permasalahan data tidak seimbang. Metode ini bekerja dengan mensintesis sampel baru dari kelas minoritas berdasarkan kemiripan jarak antar data, sehingga mampu memperbaiki distribusi data dan membantu model dalam mengenali pola dari kelas yang jarang muncul[16]. Secara umum, rumus yang digunakan untuk menghasilkan data sintesis dapat dituliskan sebagai berikut:

$$v_{syn} = v_i + (v_{knn} - v_i) \times \delta \dots \dots \dots (1)$$

Nilai acak yang hampir bernilai 0 menghasilkan data sintesis dengan karakteristik yang sangat mirip terhadap data minoritas asli. Sebaliknya, ketika nilainya mendekati 1, data sintesis tersebut cenderung memiliki kesamaan dengan sampel hasil replikasi terdekat.[17].

5. *Klasifikasi K-Nearest Neighbor*

K-NN digunakan untuk melakukan klasifikasi objek dengan menempatkannya ke dalam kelompok yang memiliki kemiripan tertinggi terhadap data latih terdekat. Penentuan kelas dilakukan melalui perhitungan jumlah K tetangga terdekat, lalu memilih kelas mayoritas. Prinsip tetangga mutual

diterapkan untuk mengatasi permasalahan outlier, dan hasilnya dipengaruhi oleh nilai K yang digunakan. Nilai K yang lebih besar biasanya mengurangi jumlah outlier yang teridentifikasi.[18].

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \dots\dots\dots(2)$$

$$d = \sqrt{\sum (a_i - b_i)^2; n_i = 1} \dots\dots\dots(3)$$

Metode ini sering diterapkan pada klasifikasi citra berkat mekanismenya yang sederhana namun andal dalam mendeteksi pola warna. Evaluasi kemiripan antar data dilakukan dengan menghitung jarak pada ruang fitur, menjadikannya sesuai digunakan saat warna menjadi ciri pembeda yang paling signifikan[19].

6. Evaluasi

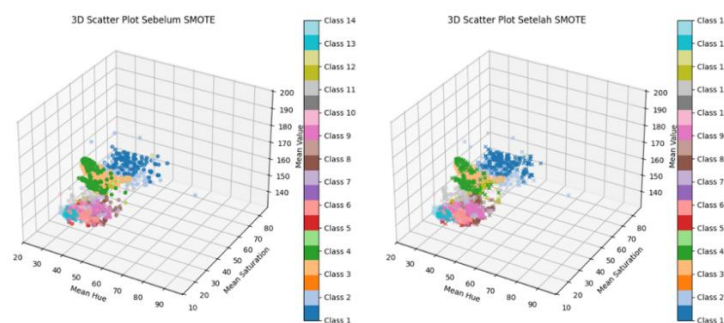
Confusion matrix merupakan representasi tabular yang membandingkan hasil prediksi model dengan kondisi aktual. Melalui matriks ini, kinerja model dalam mengidentifikasi data yang relevan serta membedakannya dari data yang tidak relevan dapat dianalisis. Selain itu, confusion matrix menjadi dasar perhitungan berbagai metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score.[20].

III. HASIL DAN PEMBAHASAN

A. Hasil *Oversampling Smote*

1. Pada dataset tidak terkondisi

Dalam tahap pengambilan citra tidak terkondisi, diperoleh dataset dengan distribusi yang timpang, di mana kelas mayor memiliki 307 sampel dan kelas minor hanya 34 sampel. Masing-masing sampel diberi label 1 hingga 14, sedangkan distribusinya ditampilkan pada Gambar 2 dalam bentuk scatter plot.



Gambar 3. Scatter Plot dataset tidak terkondisi

Gambar 3. memperlihatkan perbandingan sebaran data *dataset* tidak terkondisi sebelum dan sesudah penerapan SMOTE pada ruang fitur *Mean Hue*, *Mean Saturation*, dan *Mean Value*.

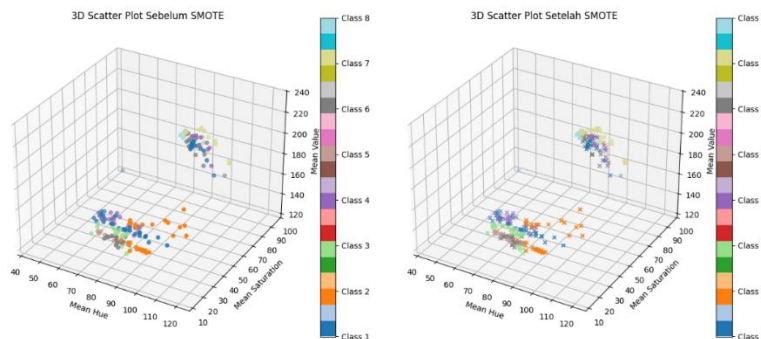
- Plot Kiri (Sebelum SMOTE)
Distribusi data menunjukkan ketidakseimbangan yang jelas, di mana beberapa kelas (misalnya Class 1, Class 2, dan Class 14) memiliki jumlah sampel jauh lebih banyak dibandingkan kelas lainnya. Beberapa kelas minoritas hanya memiliki sedikit titik data, sehingga sulit bagi model untuk mempelajari pola yang representatif.
- Plot Kanan (Sesudah SMOTE)
Setelah SMOTE diterapkan, jumlah titik data untuk kelas minoritas meningkat secara signifikan melalui penambahan sampel sintesis. Distribusi antar kelas menjadi lebih merata,

sehingga mengurangi bias model terhadap kelas mayoritas dan meningkatkan peluang klasifikasi yang lebih seimbang.

Perubahan ini memastikan setiap kelas memiliki representasi yang cukup di ruang fitur, yang pada akhirnya berdampak positif terhadap akurasi dan metrik evaluasi lainnya pada model KNN (*K-Nearest Neighbor*).

2. Hasil pada dataset terkondisi

Dataset terkondisi terdiri dari delapan kelas, dengan jumlah data pada kelas mayoritas sebanyak 44 gambar dan kelas minoritas hanya 9 gambar. Dapat dilihat pada Gambar 3.



Gambar 4. Scatter Plot Dataset Terkondisi

Gambar 4. memperlihatkan sebaran data sebelum dan sesudah SMOTE pada ruang fitur *Mean Hue*, *Mean Saturation*, dan *Mean Value*:

- Plot Kiri (Sebelum SMOTE): Distribusi data memperlihatkan adanya ketidakseimbangan, dengan beberapa kelas, seperti kelas (Class 2), memiliki jumlah sampel yang jauh lebih besar dibandingkan kelas lainnya.
- Plot Kanan (Sesudah SMOTE): Metode SMOTE menghasilkan sampel tambahan pada kelas minoritas melalui interpolasi di ruang fitur, sehingga distribusi data antar kelas menjadi lebih seimbang dan setiap kelas memiliki jumlah sampel yang memadai untuk proses pelatihan pada model KNN.

Perubahan ini diharapkan dapat meningkatkan kemampuan model dalam mengenali pola dari semua kelas, mengurangi bias terhadap kelas mayoritas, dan meningkatkan performa metrik evaluasi.

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

1. Dataset terkondisi

Pengujian pada data terkondisi dilakukan menggunakan KNN dengan nilai $k = 3, 5, 7$, dan 9 , serta skema validasi k -fold 5 dan 10 . Dataset ini memiliki kualitas pencitraan yang konsisten (latar putih, pencahayaan terkontrol).

TABEL II
HASIL UJI DATASET TERKONDISI MENGGUNAKAN K-FOLD 5

K-Fold	Nilai K	Motode Dataset	Akurasi	Precision	Recall	F1-score
5	3	Preprosesing	66.84%	76.08%	63.15%	62.74%
5	5	Preprosesing	67.36%	76.63%	59.64%	59.53%
5	7	Preprosesing	64.21%	74.02%	63.15%	63.20%
5	9	Preprosesing	66.84%	75.28%	61.40%	62.13%
5	3	Preprosesing + Smote	84.33%	90.69%	89.07%	88.86%
5	5	Preprosesing + Smote	81.05%	87.11%	84.03%	83.82%
5	7	Preprosesing + Smote	79.02%	87.58%	84.03%	83.62%
5	9	Preprosesing + Smote	77.25%	85.63%	81.51%	81.11%

TABEL III
HASIL UJI DATASET TERKONDISI MENGGUNAKAN K-FOOLD 10

K-Fold	Nilai K	Motode Dataset	Akurasi	Precision	Recall	F1-score
10	3	Preprosesing	68.42%	76.08%	63.15%	62.74%
10	5	Preprosesing	66.84%	76.63%	59.64%	59.53%
10	7	Preprosesing	65.26%	74.02%	63.15%	63.20%
10	9	Preprosesing	66.31%	75.28%	61.40%	62.13%
10	3	Preprosesing + Smote	85.32%	90.69%	89.07%	88.86%
10	5	Preprosesing + Smote	83.32%	87.11%	84.03%	83.82%
10	7	Preprosesing + Smote	81.05%	87.58%	84.03%	83.62%
10	9	Preprosesing + Smote	78.77%	85.63%	81.51%	81.11%

Dari Tabel II dan Tabel III. Pengujian pada dataset terkondisi yang memiliki kualitas pencitraan seragam, latar putih, dan pencahayaan terkontrol menunjukkan bahwa penggunaan metode K-KNN dengan bantuan SMOTE secara konsisten meningkatkan performa model. Konfigurasi terbaik diperoleh pada nilai k=3 dengan skema k-fold 10, menghasilkan akurasi 85,32%, precision 90,69%, recall 89,07%, dan F1-score 88,86%. Tren penurunan akurasi terlihat jelas seiring meningkatnya nilai k, disebabkan oleh bertambahnya jumlah tetangga dari kelas berbeda yang dipertimbangkan, sehingga peluang mis-klasifikasi meningkat, khususnya pada kelas dengan kemiripan warna tinggi.

2. Dataset Tidak Terkondisi

Pada data tidak terkondisi, kemudian diuji menggunakan algoritma KNN dengan nilai K = 3, 5, 7, dan 9, menghasilkan hasil pengujian sebagaimana ditunjukkan pada tabel berikut:

TABEL IV
HASIL UJI DATASET TIDAK TERKONDISI MENGGUNAKAN K-FOLD 5

K Fold	Nilai K	Motode Data	Akurasi	Precision	Recall	F1-score
5	3	Preprosesing	72.58%	74.59%	73.44%	72.78%
5	5	Preprosesing	72.58%	74.45%	73.44%	72.86%
5	7	Preprosesing	72.65%	74.16%	73.23%	72.43%
5	9	Preprosesing	72.39%	75.72%	74.51%	73.87%
5	3	Preprosesing + Smote	86.66%	86.41%	86.35%	86.29%
5	5	Preprosesing + Smote	84.55%	83.83%	83.48%	83.43%
5	7	Preprosesing + Smote	82.54%	81.77%	81.39%	81.35%
5	9	Preprosesing + Smote	81.59%	81.51%	81.00%	80.92%

TABEL V
HASIL UJI DATASET TIDAK TERKONDISI MENGGUNAKAN K-FOLD 10

K Fold	Nilai K	Motode Data	Akurasi	Precision	Recall	F1-score
10	3	Preprosesing	73.49%	74.59%	73.44%	72.78%
10	5	Preprosesing	74.20%	74.45%	73.44%	72.86%
10	7	Preprosesing	73.17%	74.16%	73.23%	72.43%
10	9	Preprosesing	73.62%	75.72%	74.51%	73.87%
10	3	Preprosesing + Smote	87.59%	86.41%	86.35%	86.29%
10	5	Preprosesing + Smote	85.43%	83.83%	83.48%	83.43%
10	7	Preprosesing + Smote	83.27%	81.77%	81.39%	81.35%
10	9	Preprosesing + Smote	82.41%	81.51%	81.00%	80.92%

Dilihat pada Tabel IV dan Tabel V. Hasil pada dataset tidak terkondisi, yang diambil dengan latar belakang bervariasi, pencahayaan berubah-ubah, dan posisi objek tidak seragam, menunjukkan tren yang sama. Konfigurasi terbaik juga diperoleh pada $k=3$ dengan k -fold 10, menghasilkan akurasi 87,59%, precision 86,41%, recall 86,35%, dan F1-score 86,29%. Variasi visual yang tinggi membuat tantangan klasifikasi lebih besar dibanding dataset terkondisi. SMOTE tetap mampu meningkatkan akurasi secara signifikan, sekitar 14% pada nilai $k=3$, dengan memperluas representasi kelas minoritas di ruang fitur. Meski demikian, interpolasi pada dataset dengan noise visual tinggi berisiko menghasilkan sampel sintesis yang kurang merepresentasikan kondisi nyata (sampling bias), karena KNN sangat bergantung pada jarak fitur yang dapat terdistorsi oleh perbedaan pencahayaan dan latar belakang.

IV. KESIMPULAN

Berdasarkan eksperimen, integrasi SMOTE + KNN secara konsisten meningkatkan akurasi klasifikasi pada dataset citra butterfly fish baik pada kondisi terkondisi maupun tidak terkondisi. Hasil penelitian membuktikan bahwa integrasi SMOTE dengan KNN efektif mengatasi ketidakseimbangan kelas pada data citra butterfly fish, baik dalam kondisi terkondisi maupun tidak terkondisi. Konfigurasi terbaik $K = 3$ dengan akurasi 85,32% (terkondisi), k -fold = 10 dengan akurasi 87,59% (tidak terkondisi). Peningkatan performa terjadi karena SMOTE mampu menyeimbangkan distribusi data antar kelas sehingga KNN dapat mempelajari pola secara lebih proporsional. Kontribusi utama penelitian ini adalah memberikan bukti empiris bahwa penggunaan SMOTE secara signifikan meningkatkan kinerja KNN pada klasifikasi citra dengan data imbalanced, serta mengidentifikasi bahwa nilai k kecil ($k=3$) konsisten memberikan hasil terbaik pada dua kondisi pengambilan data yang berbeda. Terlepas dari hal tersebut, penelitian ini memiliki beberapa keterbatasan, antara lain potensi overfitting akibat kemiripan tinggi antara data sintesis dan data asli, sensitivitas KNN terhadap noise visual, dan terbatasnya pengujian hanya pada satu algoritma klasifikasi. Potensi bias baru juga dapat muncul dari data sintesis yang dihasilkan SMOTE jika distribusi awal sudah bias, serta dominasi fitur tertentu (misalnya warna) yang dapat membatasi generalisasi. Untuk penelitian selanjutnya, disarankan untuk menguji kombinasi SMOTE dengan algoritma lain seperti Random Forest, XGBoost, atau CNN, menerapkan teknik feature selection atau feature extraction untuk mengurangi noise, serta menguji model pada dataset dengan variasi latar belakang dan pencahayaan yang lebih kompleks. Selain itu, hybrid resampling seperti SMOTE dengan Tomek Links dapat dipertimbangkan untuk mengurangi overlap antar kelas sehingga hasil klasifikasi menjadi lebih robust.

REFERENSI

- [1] 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.
- [2] Ridwan, E. H. Hermalian and M. Ernawati, "Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian," *Computer Science*, pp. 80-88, 2024.
- [3] A. Paliling, M. Muchtar and Fardian, "Sistem Cerdas Deteksi Kematangan Buah Naga Berbasis HSV-KNN," *Jurnal Sistem Informasi dan Teknologi Informasi*, pp. 46-55, 2025.
- [4] Yohannes, M. E. A. Rivan, S. Devella and Meiriyama, "Ekstraksi Fitur Warna Dengan Histogram HSV Untuk Klasifikasi Motif Songket Palembang," *Jurnal Teknik Informatika dan Sistem Informasi*, pp. 386-396, 2024.

- [5] 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.
- [6] D. Syahid, Jumiadi and Nursantika, "Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode K-Nearest Neighbor (Knn) Berdasarkan Nilai Hue, Saturation, Value (Hsv)," *JOIN*, p. 1, 2016.
- [7] E. Sutoyo and A. M. Fadlurrahman, "Penerapan SMOTE Untuk Mengatasi Imbalance Class Dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *Jurnal Edukasi Dan Penelitian Informatika*.
- [8] L. Ningsih and p. cholidhazia, "Klasifikasi Tingkat Kematangan Tomat Berdasarkan Warna RGB dan HSV Menggunakan Algoritma KNN" [Classification of Tomato Maturity Levels Based on RGB and HSV Color Using KNN Algorithm]," *Jurnal of Artificial intelligence and digital business*, pp. 25-30, 2022.
- [9] 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.
- [10] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan K-Nearest Neighbor," *Journal Information System Development*, pp. 44-49, 2018.
- [11] D. Agustin, R. D. Atmaja, S.T, M,T and Azizah, S.T, M.T, "Pengolahan Citra Digital Untuk Mengklasifikasi Golongan Kendaraan Dengan Metode Parameter Dasar Geometrik," *E Proceeding Of Engineering*, vol. 4, no. 1, pp. 115-123, 2017.
- [12] 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.
- [13] F. D. Astuti and f. N. lenti, "Implementasi SMOTE Untuk Mengatasi Imbalance Class Pada Klasifikasi Car Evolution Menggunakan KNN," *Informatika STMIK AKAKOM, Yogyakarta*, vol. 13, no. 1, pp. 89-98, 2021.
- [14] A. Fauziah and J. Hernadi, "Klasifikasi Data Tak Seimbang Menggunakan Algoritma Random Forest dengan SMOTE dan SMOTE-ENN (Studi Kasus pada Data Stunting)," *Jurnal Informatika dan Komputer*, vol. 17, no. 2, pp. 38-47, 2024.
- [15] c. Y. Jerandu, p. batarius and a. a. j. sinlae, "Identifikasi Kualitas Kesegaran Ikan Menggunakan Algoritma K-Nearest Neighbor Berdasarkan Ekstraksi Ciri Warna Hue, Saturation, Dan Value (HSV)," *Building of Informatic Technology And Science (BITS)*, vol. 4, no. 3, pp. 1536-1547, 2022.
- [16] 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.
- [17] 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.
- [18] 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.