

Pendekatan *Machine Learning* Untuk Prediksi Motif Batik Indonesia

Wahyu Marzian Putra^{1*}, Erliyan Redy Susanto², Neneng³

^{1,2,3}Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

E-mail: ^{1*}wahyu.marzian.putra.mhs@teknokrat.ac.id, ²erliyan.redy@teknokrat.ac.id,

³neneng@teknokrat.ac.id

Abstrak

Batik merupakan warisan budaya tak benda Indonesia yang diakui UNESCO sejak tahun 2009. Perkembangan teknologi kecerdasan buatan, khususnya *deep learning*, membuka peluang otomatisasi klasifikasi visual batik. Namun, klasifikasi otomatis batik Indonesia masih menghadapi tantangan utama, yaitu ketidakseimbangan kelas pada dataset (*imbalanced dataset*) dan kemiripan visual antarkelas yang tinggi. Penelitian ini mengembangkan sistem klasifikasi 20 jenis batik Indonesia menggunakan model YOLOv11 yang diintegrasikan dengan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) serta skema *5-Fold Stratified Cross-Validation* guna memperoleh evaluasi yang objektif. Dataset terdiri atas 983 gambar yang diakuisisi dari repositori Kaggle. SMOTE diterapkan pada representasi fitur piksel berdimensi tinggi ($d = 12,288$) untuk menyeimbangkan distribusi antarkelas tanpa mengorbankan informasi kelas mayoritas. Model YOLOv11 dilatih pada setiap lipatan dengan konfigurasi *hyperparameter* terkontrol menggunakan pengoptimalan AdamW. Hasil penelitian menunjukkan nilai *mean Average Precision* (mAP) keseluruhan sebesar 0,8251 dan rata-rata mAP per lipatan sebesar $0,8788 \pm 0,0264$, dengan konsistensi metrik yang tinggi di seluruh lipatan validasi. Analisis matriks kebingungan mengindikasikan bahwa *misclassification* cenderung terjadi antara kelas batik yang memiliki kemiripan visual tinggi. Kerangka kerja yang diusulkan berpotensi diimplementasikan sebagai sistem identifikasi batik otomatis untuk mendukung pelestarian warisan budaya, autentifikasi produk pada platform *e-commerce*, serta dokumentasi koleksi museum di era digital.

Kata Kunci: Klasifikasi Gambar, YOLOv11, SMOTE, Cross-Validation, Deep Learning

Abstract

Batik is an Indonesian intangible cultural heritage recognized by UNESCO since 2009. The development of artificial intelligence technology, especially deep learning, opens up opportunities for automated visual classification of batik. However, automatic classification of Indonesian batik still faces major challenges, namely class imbalance in the dataset (*imbalanced dataset*) and high visual similarity between classes. This study develops a classification system for 20 types of Indonesian batik using the YOLOv11 model integrated with the Synthetic Minority Over-sampling Technique (SMOTE) and the 5-Fold Stratified Cross-Validation scheme to obtain objective evaluation. The dataset consists of 983 images acquired from the Kaggle repository. SMOTE is applied to the high-dimensional pixel feature representation ($d = 12,288$) to balance the distribution between classes without sacrificing the majority class information. The YOLOv11 model is trained on each fold with a controlled hyperparameter configuration using AdamW optimization. The results showed an overall mean Average Precision (mAP) of 0.8251 and an average mAP per fold of 0.8788 ± 0.0264 , with high metric consistency across validation folds. Confusion matrix analysis indicated that misclassification tends to occur between batik classes with high visual similarity. The proposed framework has the potential to be implemented as an automatic batik identification system to support cultural heritage preservation, product authentication on e-commerce platforms, and museum collection documentation in the digital era.

Keyword: Image Classification, YOLOv11, SMOTE, Cross-Validation, Deep Learning

*Corresponding Author:

Wahyu Marzian Putra, Universitas Teknokrat Indonesia, Indonesia

Email: wahyu.marzian.putra.mhs@teknokrat.ac.id

1. PENDAHULUAN

Batik merupakan warisan budaya tak benda Indonesia yang telah diakui secara internasional oleh UNESCO sejak tahun 2009, menjadikannya salah satu identitas kebudayaan bangsa yang memiliki nilai estetika, historis, dan ekonomi yang tinggi[1]. Keberagaman motif batik di Indonesia sangat luas, tersebar dari Sabang hingga Merauke, dengan setiap daerah memiliki ciri khas dan filosofi tersendiri dalam pola, warna, serta teknik pembuatannya. Dalam dekade terakhir, industri batik mengalami pertumbuhan yang signifikan seiring dengan meningkatnya kesadaran masyarakat terhadap pelestarian budaya serta berkembangnya platform perdagangan digital yang memfasilitasi distribusi batik secara nasional maupun internasional. Fenomena digitalisasi ini mendorong kebutuhan akan sistem identifikasi otomatis terhadap jenis-jenis batik, mengingat jumlah motif yang sangat beragam dan sering kali sulit dibedakan oleh masyarakat awam maupun pelaku industri. Perkembangan teknologi kecerdasan buatan, khususnya di bidang *computer vision* dan pembelajaran mendalam (*deep learning*), telah membuka peluang baru untuk mengotomatisasi proses klasifikasi visual batik dengan tingkat akurasi yang tinggi dan konsisten [2], [3], [4].

Meskipun demikian, klasifikasi otomatis jenis batik Indonesia masih menghadapi sejumlah tantangan substansial yang belum sepenuhnya teratasi. Pertama, dataset batik yang tersedia secara umum cenderung bersifat *imbalanced* atau tidak seimbang, di mana jumlah sampel pada setiap kelas jenis batik bervariasi secara signifikan, sehingga model pembelajaran mesin cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas [5]. Kedua, visual similarity antarjenis batik yang tinggi – misalnya antara batik Pekalongan, batik Solo, dan batik Cirebon yang memiliki kemiripan pola geometris dan palet warna – menambah kompleksitas dalam proses ekstraksi dan pembelajaran fitur diskriminatif. Ketiga, keterbatasan jumlah sampel per kelas dalam dataset yang umum digunakan dalam penelitian serupa menjadi hambatan dalam melatih model yang mampu menggeneralisasi dengan baik. Oleh karena itu, diperlukan pendekatan yang tidak hanya mengandalkan arsitektur model yang kuat, tetapi juga memperhatikan strategi penanganan ketidakseimbangan data serta skema evaluasi yang komprehensif guna memastikan keandalan dan reproduktibilitas hasil klasifikasi [6].

Sejumlah penelitian terdahulu telah mengupayakan klasifikasi otomatis motif batik menggunakan berbagai pendekatan deep learning. Rasyidi dan Bariyah (2020) menggunakan Convolutional Neural Network (CNN) dengan arsitektur DenseNet untuk mengklasifikasikan enam motif batik dan berhasil mencapai akurasi 94% [6]. Wijaya dan Suharjito (2021) menerapkan VGG-16 pada 10 kelas batik dengan skema 10-Fold Cross-Validation dan memperoleh akurasi 88,5% [21]. Nugroho et al. (2022) menggunakan EfficientNet-B0 dengan teknik random oversampling pada 12 kelas batik dan mencatat akurasi 93,1% [22]. Saputra et al. (2023) memanfaatkan MobileNetV2 berbasis transfer learning untuk mengklasifikasikan 8 motif batik dengan akurasi 89,7% [23]. Meskipun demikian, studi-studi tersebut memiliki beberapa keterbatasan yang menjadi celah penelitian (*research gap*): jumlah kelas batik yang diklasifikasikan masih relatif sedikit (maksimal 15 kelas), sebagian besar tidak menangani ketidakseimbangan kelas secara sistematis, dan penggunaan train/test split tunggal rentan menghasilkan estimasi kinerja yang bias pada dataset kecil. Selain itu, belum ada studi yang memanfaatkan arsitektur YOLOv11 yang dirancang khusus untuk tugas visual hierarkis dalam konteks klasifikasi batik Indonesia.

Sejumlah penelitian terdahulu telah mengupayakan klasifikasi otomatis motif batik menggunakan berbagai pendekatan deep learning. Rasyidi dan Bariyah (2020) menggunakan Convolutional Neural Network (CNN) dengan arsitektur DenseNet untuk mengklasifikasikan enam motif batik dan berhasil mencapai akurasi 94% [6]. Wijaya dan Suharjito (2021) menerapkan VGG-16 pada 10 kelas batik dengan skema 10-Fold Cross-Validation dan memperoleh akurasi 88,5% [21]. Nugroho et al. (2022) menggunakan EfficientNet-B0 dengan teknik random oversampling pada 12 kelas batik dan mencatat akurasi 93,1% [22]. Saputra et al. (2023) memanfaatkan MobileNetV2 berbasis transfer learning untuk mengklasifikasikan 8 motif batik dengan akurasi 89,7% [23]. Meskipun demikian, studi-studi tersebut memiliki beberapa keterbatasan yang menjadi celah penelitian (*research gap*): jumlah kelas batik yang diklasifikasikan masih relatif sedikit (maksimal 15 kelas), sebagian besar tidak menangani ketidakseimbangan kelas secara sistematis, dan penggunaan train/test split tunggal rentan menghasilkan estimasi kinerja yang bias pada dataset kecil. Selain itu, belum ada studi yang

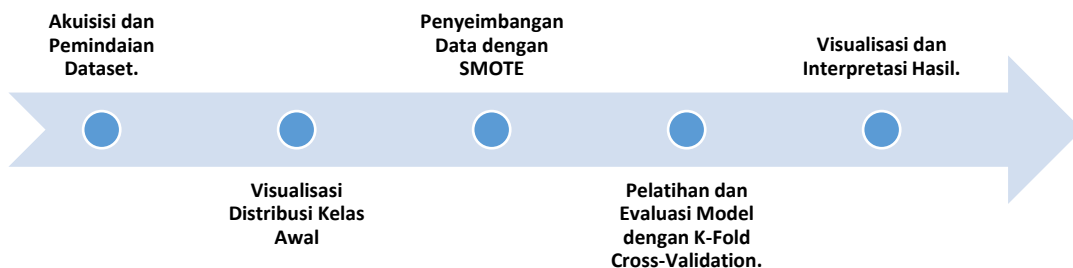
memanfaatkan arsitektur YOLOv11 yang dirancang khusus untuk tugas visual hierarkis dalam konteks klasifikasi batik Indonesia.

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem klasifikasi 20 jenis batik Indonesia menggunakan model YOLOv11 (You Only Look Once version 11) yang dipadukan dengan teknik Synthetic Minority Over-sampling Technique (SMOTE) untuk menangani ketidakseimbangan kelas, serta skema 5-Fold Stratified Cross-Validation untuk menjamin evaluasi yang objektif dan komprehensif. Tahapan penelitian meliputi ekstraksi dan pemindaian dataset batik yang terdiri atas 983 gambar pada 20 kelas, dilanjutkan dengan penerapan SMOTE pada representasi fitur piksel untuk menghasilkan sampel sintesis guna menyeimbangkan distribusi kelas, kemudian pelatihan model klasifikasi YOLOv11 pada masing-masing lipatan (fold) dengan konfigurasi hyperparameter yang terkontrol, serta evaluasi kinerja model menggunakan metrik accuracy, recall, precision, F1-score, dan mean Average Precision (mAP) baik secara makro maupun per kelas. YOLOv11 dipilih bukan sekadar sebagai model deteksi objek, melainkan karena modul klasifikasi gambarnya memanfaatkan backbone konvolusional dengan mekanisme perhatian (attention) dan fusi fitur multi-skala, sehingga mampu menangkap detail lokal tekstur batik sekaligus pola global komposisi motif secara simultan – keunggulan yang tidak dimiliki model klasifikasi murni seperti ResNet atau VGG yang dirancang khusus untuk ImageNet tanpa optimasi fitur hierarkis visual semacam ini. Arsitektur YOLOv11 yang ringan (varian nano) juga memungkinkan pelatihan efisien pada dataset berukuran terbatas. Sementara itu, SMOTE diterapkan untuk mengatasi permasalahan ketidakseimbangan data tanpa mengorbankan informasi dari kelas mayoritas.

2. METODE PENELITIAN

2.1 Tahapan Studi

Penelitian ini dilaksanakan melalui lima tahapan utama yang tersusun secara sistematis dan berkesinambungan, sebagaimana diilustrasikan pada Gambar 1. Setiap tahapan dirancang untuk memastikan reproduktibilitas, validitas, dan keandalan hasil klasifikasi yang dihasilkan. Tahapan tersebut adalah sebagai berikut.



Gambar 1. Tahapan Penelitian

Tahap I – Akuisisi dan Pemindaian Dataset. Pada tahap awal, dataset gambar batik Indonesia diakuisisi dari repositori daring Kaggle. Dataset kemudian diekstrak, dipindai, dan diverifikasi untuk memastikan integritas setiap berkas gambar. Proses ini meliputi inspeksi format berkas, dimensi piksel, dan konsistensi penamaan direktori kelas. Selain itu, dilakukan penghitungan distribusi awal (*initial distribution*) setiap kelas untuk mengidentifikasi ketidakseimbangan yang ada.

Tahap II – Visualisasi Distribusi Kelas Awal. Distribusi kelas divisualisasikan dalam bentuk diagram batang dan diagram lingkaran guna memberikan gambaran komprehensif mengenai proporsi sampel pada masing-masing kelas. Visualisasi ini menjadi dasar pengambilan keputusan dalam menentukan strategi penyeimbangan data pada tahap berikutnya.

Tahap III – Penyeimbangan Data dengan SMOTE. Karena dataset bersifat *imbalanced*, diterapkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) pada representasi fitur piksel yang telah

dinormalisasi. Hasil SMOTE divisualisasikan untuk membandingkan distribusi kelas sebelum dan sesudah penyeimbangan [7].

Tahap IV – Pelatihan dan Evaluasi Model dengan K-Fold Cross-Validation. Dataset hasil SMOTE dilatih dan dievaluasi menggunakan model YOLOv11 dalam skema *5-Fold Stratified Cross-Validation*. Setiap lipatan (*fold*) menghasilkan metrik evaluasi yang terekam secara independen.

Tahap V – Visualisasi dan Interpretasi Hasil. Hasil akhir divisualisasikan melalui matriks kebingungan (*confusion matrix*), grafik metrik per kelas, grafik metrik per lipatan, serta grafik *mean Average Precision* (mAP). Laporan akhir disusun sebagai rangkuman kuantitatif keseluruhan kinerja sistem.

2.2 Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan dataset gambar batik Indonesia yang diperoleh dari repositori Kaggle. Dataset ini terdiri atas **20 kelas** yang mewakili motif batik dari berbagai daerah di Indonesia, dengan total **983 gambar** dalam format JPEG. Rincian kelas-kelas tersebut beserta asal daerahnya disajikan pada Tabel 1.

Tabel 1. Daftar Kelas dalam Dataset Batik Indonesia

No.	Kelas	Asal Daerah	No.	Kelas	Asal Daerah
1	Batik Bali	Bali	11	Batik Lasem	Jawa Tengah
2	Batik Betawi	DKI Jakarta	12	Batik Megamendung	Jawa Barat
3	Batik Celup	Jawa Timur	13	Batik Parang	Yogyakarta
4	Batik Cendrawasih	Papua	14	Batik Pekalongan	Jawa Tengah
5	Batik Ceplok	Jawa Tengah	15	Batik Priangan	Jawa Barat
6	Batik Ciamis	Jawa Barat	16	Batik Sekar	Jawa Tengah
7	Batik Garutan	Jawa Barat	17	Batik Sidoluhur	Jawa Tengah
8	Batik Gentongan	Jawa Timur	18	Batik Sidomukti	Jawa Tengah
9	Batik Kawung	Yogyakarta	19	Batik Sogan	Jawa Tengah
10	Batik Keraton	Yogyakarta	20	Batik Tambal	Jawa Tengah

Setiap kelas memiliki jumlah sampel berkisar antara 45 hingga 50 gambar per kelas, sehingga total keseluruhan berjumlah 983 gambar. Gambar-gambar tersebut disusun dalam struktur direktori hierarkis, di mana setiap subdirektori mewakili satu kelas batik. Kondisi ketidakseimbangan distribusi kelas inilah yang mendorong penerapan teknik oversampling SMOTE pada tahap selanjutnya.

2.3 Metode yang Digunakan

2.3.1 Ekstraksi Fitur dan Normalisasi

Sebelum penerapan SMOTE, setiap gambar dalam dataset direpresentasikan dalam bentuk vektor fitur numerik. Proses ekstraksi fitur dilakukan dengan meresize setiap gambar berukuran asli ke dimensi target $F \times F \times 3$ piksel, di mana $F = 64$, kemudian menormalisasi nilai piksel ke rentang $[0,1]$. Secara matematis, representasi fitur suatu gambar I_i didefinisikan sebagai:

$$X_i = \text{flatten} \left(\frac{I_i^{\text{resized}}}{255} \right) \in \mathbb{R}^d, d = F \times F \times 3 \quad (1)$$

di mana I_i^{resized} adalah gambar i yang telah di-*resize*, operasi *flatten* (.) mengubah tensor tiga dimensi menjadi vektor satu dimensi, dan d adalah dimensi ruang fitur. Dengan $F = 64$, maka $d = 64 \times 64 \times 3 = 12.288$ dimensi.

2.3.2 Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE adalah teknik oversampling sintesis yang diperkenalkan oleh Chawla *et al.* (2002) untuk menangani permasalahan ketidakseimbangan kelas pada dataset pembelajaran mesin. Berbeda dengan oversampling konvensional yang sekadar menduplikasi sampel minoritas, SMOTE menghasilkan sampel sintesis baru melalui interpolasi linier antara sampel minoritas yang ada dan tetangga terdekatnya dalam ruang fitur [8].

Prosedur SMOTE dilaksanakan sebagai berikut:

Langkah 1. Untuk setiap sampel minoritas x_i dalam kelas minoritas c , cari k tetangga terdekatnya dalam ruang fitur menggunakan jarak Euclidean. Jarak Euclidean antara dua vektor fitur x_i dan x_j didefinisikan sebagai:

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^d (x_{i,m} - x_{j,m})^2} \quad (2)$$

di mana d adalah dimensi vektor fitur, dan $x_{i,m}$ serta $x_{j,m}$ adalah komponen ke- m dari vektor x_i dan x_j secara berturut-turut.

Langkah 2. Pilih secara acak salah satu dari k tetangga terdekat tersebut, misalkan x_{nn} .

Langkah 3. Hitung vektor selisih antara x_i dan tetangga terdekatnya:

$$\delta = x_{nn} - x_i \quad (3)$$

Langkah 4. Hasilkan sampel sintetis baru x_{syn} dengan mengalikan vektor selisih dengan bilangan acak $\lambda \in [0,1]$:

$$x_{syn} = x_i + \lambda \cdot \delta \quad (4)$$

atau secara komponen-komponen:

$$x_{syn,m} = x_{i,m} + \lambda \cdot (x_{nn,m} - x_{i,m}), \quad m = 1, 2, \dots, d \quad (5)$$

Langkah 5. Ulangi Langkah 2-4 hingga jumlah sampel pada kelas minoritas mencapai jumlah yang diinginkan (biasanya menyamai jumlah kelas mayoritas).

Dalam penelitian ini, parameter k (tetangga terdekat) ditetapkan sebesar $k = 4$, dengan mempertimbangkan bahwa jumlah minimum sampel per kelas pada dataset adalah sekitar 45, sehingga $k = 4$ memenuhi syarat $k < n_{min}$, di mana n_{min} adalah jumlah sampel kelas minoritas.

2.3.3 Model Klasifikasi YOLOv11

YOLOv11 (*You Only Look Once version 11*) merupakan iterasi terbaru dari keluarga arsitektur YOLO yang dikembangkan oleh Ultralytics. Meskipun YOLO secara historis dikenal sebagai model deteksi objek, versi-versi terbaru YOLO menyediakan modul klasifikasi gambar (*image classification*) yang memanfaatkan *backbone* konvolusional untuk ekstraksi fitur hierarkis, diikuti oleh lapisan klasifikasi untuk memetakan representasi fitur ke distribusi probabilitas kelas [9]

Arsitektur YOLOv11 Classifier secara umum terdiri atas tiga komponen utama:

(a) Backbone (Jaringan Konvolusional). Backbone bertugas mengekstraksi fitur hierarkis dari citra input. Pada setiap lapisan konvolusi, operasi konvolusi diterapkan sebagai berikut:

$$Y^{(l)} = f \left(\sum_{c=1}^C W_c^{(l)} * X_c^{(l-1)} + b^{(l)} \right) \quad (6)$$

di mana $Y^{(l)}$ adalah peta fitur output pada lapisan ke- l , $W_c^{(l)}$ adalah kernel konvolusi untuk saluran input ke- c , $X_c^{(l-1)}$ adalah peta fitur input dari lapisan sebelumnya, $b^{(l)}$ adalah bias, $*$ menunjukkan operasi konvolusi diskrit, dan $f(\cdot)$ adalah fungsi aktivasi (misalnya SiLU/Swish).

(b) Neck. Komponen *neck* berfungsi sebagai penghubung antara *backbone* dan *head*, yang bertujuan untuk menggabungkan dan memperhalus representasi fitur dari berbagai level resolusi.

(c) Head (Klasifikasi). Pada modul klasifikasi, representasi fitur dari *backbone* diproyeksikan ke ruang kelas melalui lapisan *fully connected* (dense) dan fungsi aktivasi softmax. Probabilitas prediksi untuk kelas ke- j didefinisikan sebagai:

$$P(y = j | x) = \frac{e^{z_j}}{\sum_{c=1}^C e^{z_c}} \quad (7)$$

di mana z_j adalah *logit* (skor mentah) untuk kelas ke- j , C adalah jumlah total kelas (dalam penelitian ini $C = 20$), dan $P(y = j | x)$ adalah probabilitas bersyarat bahwa input x termasuk ke dalam kelas j .

Fungsi Loss. Pelatihan model menggunakan fungsi *Cross-Entropy Loss*, yang didefinisikan sebagai:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{i,j} \cdot \log(\hat{y}_{i,j}) \quad (8)$$

di mana N adalah jumlah sampel dalam satu *batch*, $y_{i,j}$ adalah *one-hot encoded* label sebenarnya untuk sampel ke- i dan kelas ke- j , dan $\hat{y}_{i,j} = P(y = j | x_i)$ adalah probabilitas prediksi model [10].

Optimisasi. Model dioptimasi menggunakan algoritma *AdamW* (*Adam with Weight Decay*), yang memperbarui parameter model sebagai berikut:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \end{aligned}$$

$$\vartheta_{t+1} = \vartheta_t - n \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} + \lambda \theta_t \right) \quad (9)$$

di mana g_t adalah gradien pada waktu ke- t , m_t dan v_t adalah estimasi momen pertama dan kedua, β_1 dan β_2 adalah faktor peluruhan (*decay rates*), n adalah *learning rate*, ϵ adalah konstanta kecil untuk stabilitas numerik, dan λ adalah koefisien *weight decay* [11], [12].

2.3.4 K-Fold Stratified Cross-Validation

Untuk memastikan evaluasi model yang objektif dan mengurangi bias yang mungkin timbul dari pembagian data tunggal, penelitian ini menerapkan skema *K-Fold Stratified Cross-Validation* dengan $K = 5$. Pada skema ini, dataset dibagi menjadi K lipatan (*fold*) yang saling eksklusif, dengan proporsi kelas pada setiap lipatan dijaga agar serupa dengan proporsi kelas pada dataset keseluruhan [13].

Prosedur K-Fold Stratified Cross-Validation:

Langkah 1. Dataset $D = \{(x_i, y_i)\}_{i=1}^N$ dipartisi menjadi K subset $D_1, D_2, \dots, D_K$ dengan ukuran yang kira-kira sama, yaitu $|D_K| \approx N/K$, dan mempertahankan proporsi kelas:

$$\frac{|\{(x,y) \in D_k : y=c\}|}{|D_k|} \approx \frac{|\{(x,y) \in D : y=c\}|}{|N|} \quad (10)$$

Langkah 2. Untuk setiap iterasi ke- k (dengan $k = 1, 2, \dots, K$), subset D_k digunakan sebagai data validasi, sedangkan $K - 1$ subset sisanya digunakan sebagai data pelatihan:

$$D_{train}^{(k)} = U_{j \neq k} D_j, D_{val}^{(k)} = D_k \quad (11)$$

Langkah 3. Model dilatih pada $D_{train}^{(k)}$ dan dievaluasi pada $D_{val}^{(k)}$ untuk menghasilkan metrik evaluasi $M^{(k)}$.

Langkah 4. Metrik keseluruhan dihitung sebagai rata-rata dari seluruh lipatan:

$$M_{mean} = \frac{1}{K} \sum_{k=1}^K M^{(k)}, \quad M_{std} = \sqrt{\frac{1}{K} \sum_{k=1}^K (M^{(k)} - M_{mean})^2} \quad [14]. \quad (12)$$

2.3.5 Metrik Evaluasi

Kinerja model dievaluasi menggunakan lima metrik utama, baik secara makro (*macro-averaged*) maupun per kelas (*per-class*).

(a) **Accuracy.** Proporsi prediksi yang benar dari seluruh sampel:

$$\frac{1}{N} \sum_{i=1}^N 1(\hat{y}_i = y_i) \quad (13)$$

di mana $1(\cdot)$ adalah fungsi indikator yang bernilai 1 jika kondisi terpenuhi, dan 0 jika sebaliknya.

(b) **Recall (Sensitivitas) Makro.** Rata-rata *recall* dari seluruh kelas:

$$\text{Recall}_{\text{makro}} = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FN_c} \quad (14)$$

di mana TP_c (*True hierarkis Positive*) adalah jumlah sampel kelas C yang diklasifikasikan dengan benar, dan FN_c (*False Negative*) adalah jumlah sampel kelas c yang salah diklasifikasikan ke kelas lain.

(c) **Precision Makro.** Rata-rata *precision* dari seluruh kelas:

$$\text{Precision}_{\text{makro}} = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FP_c} \quad (15)$$

di mana FP_c (*False Positive*) adalah jumlah sampel yang sebenarnya bukan kelas c tetapi diklasifikasikan sebagai kelas c .

(d) **F1-Score Makro.** Rata-rata harmonik dari *precision* dan *recall*:

$$F1_{\text{makro}} = \frac{1}{C} \sum_{c=1}^C \frac{2 \cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad (16)$$

(e) **Mean Average Precision (mAP).** Metrik ini mengukur kualitas ranking probabilitas prediksi model. Untuk setiap kelas c , *Average Precision* (AP) dihitung sebagai luas di bawah kurva *Precision-Recall*:

$$AP_c = \sum_n (R_n - R_{n-1}) \cdot P_n \quad (17)$$

di mana P_n dan R_n adalah *precision* dan *recall* pada *threshold* ke- n . Selanjutnya, mAP dihitung sebagai rata-rata AP seluruh kelas:

$$\text{mAP} = \frac{1}{C} \sum_{c=1}^C AP_c \quad [15], [16] \quad (18)$$

2.3.6 Konfigurasi Eksperimen

Konfigurasi *hyperparameter* yang digunakan dalam seluruh eksperimen dirangkum pada Tabel 2.

Tabel 2. Konfigurasi Hyperparameter Eksperimen

Parameter	Nilai	Keterangan
Model	YOLOv11 Nano Classifier	Arsitektur ringan untuk klasifikasi
Ukuran Input	224 x 224 piksel	Dimensi citra input model
Ukuran Fitur SMOTE	64 x 64 x 3	Dimensi untuk ekstraksi fitur
Epoch	30	Jumlah iterasi pelatihan per fold
Batch Size	16	Ukuran batch pelatihan
Optimizer	AdamW	Algoritma optimisasi
Learning Rate	1×10^{-3}	Laju pembelajaran awal
Weight Decay	5×10^{-4}	Regularisasi L2
Jumlah Fold	5	Lipatan cross-validation
SMOTE k-neighbors	4	Tetangga terdekat untuk SMOTE
Early Stopping	10 epoch	Patience untuk early stopping
Augmentasi	Ya	Augmentasi bawaan YOLOv11[17]
Random State	42	Seed untuk reproduktibilitas

Seluruh implementasi dilakukan menggunakan bahasa pemrograman Python 3 dengan pustaka utama meliputi *Ultralytics* (YOLOv11), *imbalanced-learn* (SMOTE), *scikit-learn* (evaluasi dan cross-validation), *matplotlib* dan *seaborn* (visualisasi), serta *PyTorch* sebagai *backend* komputasi [18].

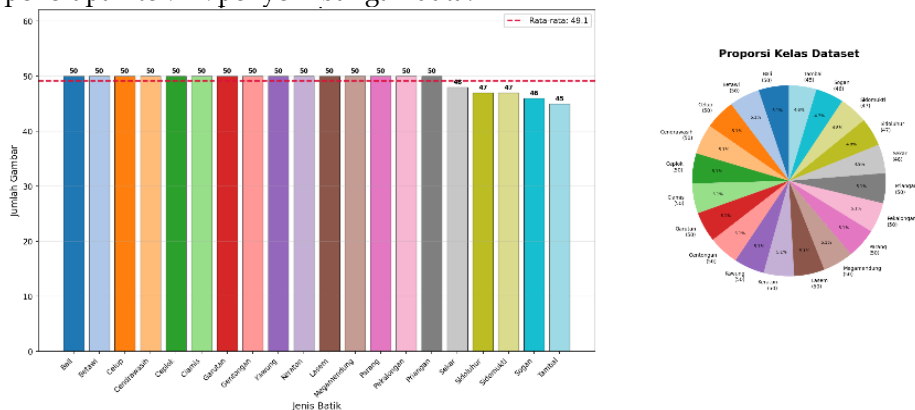
3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Hasil penelitian menunjukkan bahwa integrasi YOLOv11 dengan SMOTE dan K-Fold Cross-Validation mampu menghasilkan sistem klasifikasi yang andal dalam mengenali 20 jenis batik Indonesia, dengan metrik evaluasi yang konsisten di seluruh lipatan validasi. Penelitian ini memberikan manfaat ganda, baik secara akademis maupun praktis: secara akademis, studi ini berkontribusi pada pengembangan literatur di bidang klasifikasi gambar tekstil tradisional dengan menyediakan kerangka kerja (framework) yang terintegrasi antara penyeimbangan data, arsitektur deteksi mutakhir, dan protokol evaluasi yang ketat; secara praktis, sistem ini berpotensi diimplementasikan sebagai alat bantu identifikasi batik bagi pelaku industri, kurator museum, maupun platform e-commerce untuk memastikan autentifikasi dan klasifikasi jenis batik secara otomatis, cepat, dan akurat, sehingga turut mendukung pelestarian dan dokumentasi warisan budaya Indonesia di era digital.

3.1.1 Distribusi Kelas Dataset Sebelum SMOTE

Distribusi kelas pada dataset batik Indonesia sebelum penerapan SMOTE divisualisasikan pada Gambar 1. Terlihat bahwa dataset terdiri atas 20 kelas batik dengan jumlah sampel yang bervariasi antarkelas. Sebagian besar kelas memiliki jumlah sampel berkisar antara 45 hingga 50 gambar, dengan total keseluruhan sebanyak 983 gambar. Proporsi setiap kelas terhadap total dataset terlihat relatif seragam melalui diagram lingkaran, meskipun terdapat sedikit ketidakseimbangan antar kelas yang menjadi dasar penerapan teknik penyeimbangan data.

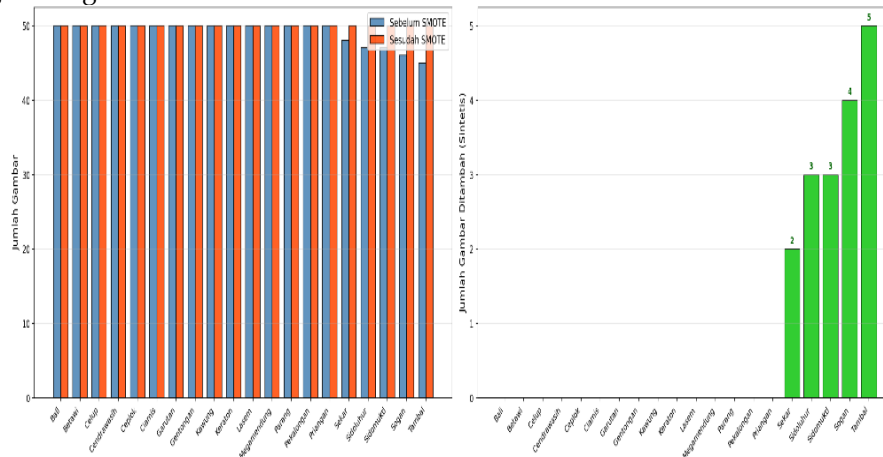


Gambar 1. Distribusi Kelas Dataset Batik Indonesia (Sebelum SMOTE)

Berdasarkan diagram batang pada Gambar 1, garis putus-putus merah menunjukkan rata-rata jumlah gambar per kelas yang digunakan sebagai acuan untuk mengidentifikasi kelas-kelas dengan jumlah sampel di bawah maupun di atas rata-rata. Meskipun ketidakseimbangan yang tergolong ringan, penerapan SMOTE tetap diperlukan untuk memastikan bahwa model tidak mengalami bias terhadap kelas-kelas tertentu selama proses pelatihan.

3.1.2 Dampak Penyeimbangan Data dengan SMOTE

Gambar 2 menampilkan perbandingan distribusi kelas sebelum dan sesudah penerapan SMOTE, serta jumlah gambar sintetis yang ditambahkan pada setiap kelas. Panel kiri menunjukkan perbandingan jumlah gambar per kelas antara kondisi awal (biru) dan kondisi setelah SMOTE (merah), sedangkan panel kanan menggambarkan delta atau selisih jumlah gambar sintetis yang ditambahkan pada masing-masing kelas.

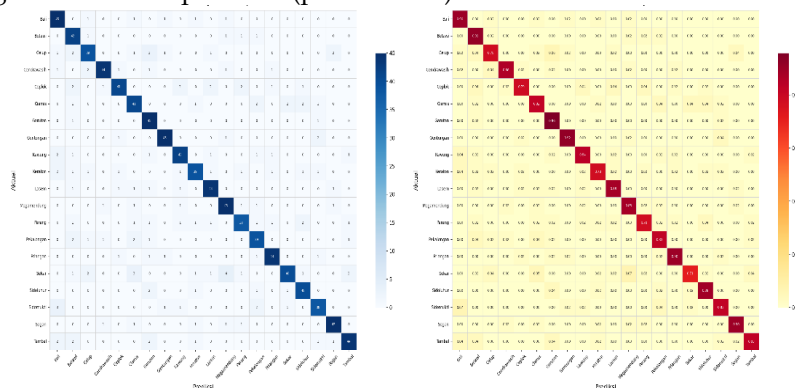


Gambar 2. Perbandingan Distribusi Kelas — Sebelum vs Sesudah SMOTE

Hasil SMOTE menunjukkan bahwa seluruh kelas berhasil diseimbangkan hingga mencapai jumlah sampel yang seragam. Gambar-gambar sintetis dihasilkan melalui interpolasi linier antara sampel minoritas dan tetangga terdekatnya dalam ruang fitur berdimensi $d = 12.288$ ($\$64 \times 64 \times 3$). Panel kanan pada Gambar 2 memperlihatkan bahwa kelas-kelas dengan jumlah sampel awal yang lebih sedikit menerima lebih banyak gambar sintetis dibandingkan kelas yang sudah memiliki jumlah sampel mendekati nilai maksimum. Proses ini menghasilkan dataset yang seimbang secara distribusi tanpa mengorbankan informasi dari kelas mayoritas, sehingga model dapat belajar representasi fitur yang lebih adil untuk seluruh kelas batik.

3.1.3 Matriks Kebingungan (*Confusion Matrix*)

Matriks kebingungan yang dihitung secara agregat dari seluruh lipatan validasi silang ditampilkan pada Gambar 3, yang mencakup matriks dalam bentuk jumlah sampel mentah (panel kiri) dan matriks yang dinormalisasi per baris (panel kanan).

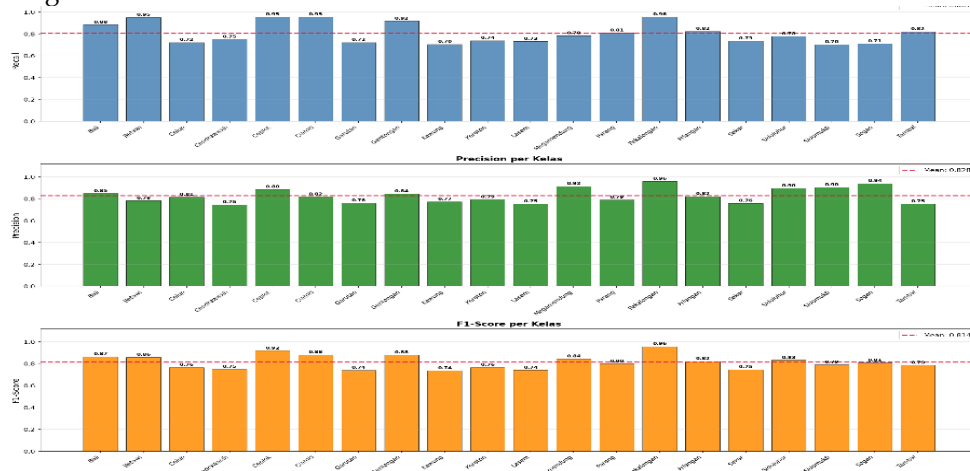


Gambar 3. Confusion Matrix — Klasifikasi Batik Indonesia (5-Fold CV, YOLOv11, 20 Kelas)

Matriks kebingungan yang dinormalisasi pada panel kanan Gambar 3 menunjukkan bahwa sebagian besar kelas memiliki nilai diagonal yang mendekati 1.0, yang mengindikasikan bahwa model mampu mengklasifikasikan sampel ke kelas yang benar dengan tingkat akurasi yang tinggi. Nilai-nilai di luar diagonal yang tinggi pada beberapa sel tertentu menunjukkan adanya kebingungan (*misclassification*) antar kelas yang memiliki kemiripan visual, seperti antara batik-batik yang berasal dari wilayah geografis yang berdekatan atau memiliki pola geometris serupa. Temuan ini konsisten dengan tantangan yang diidentifikasi dalam studi pendahuluan mengenai *visual similarity* antarjenis batik Indonesia.

3.1.4 Metrik Evaluasi Per Kelas

Gambar 4 menyajikan metrik evaluasi per kelas secara terpisah, meliputi *Recall*, *Precision*, dan *F1-Score* untuk masing-masing dari 20 kelas batik, yang dihitung sebagai agregat dari seluruh lipatan validasi silang.

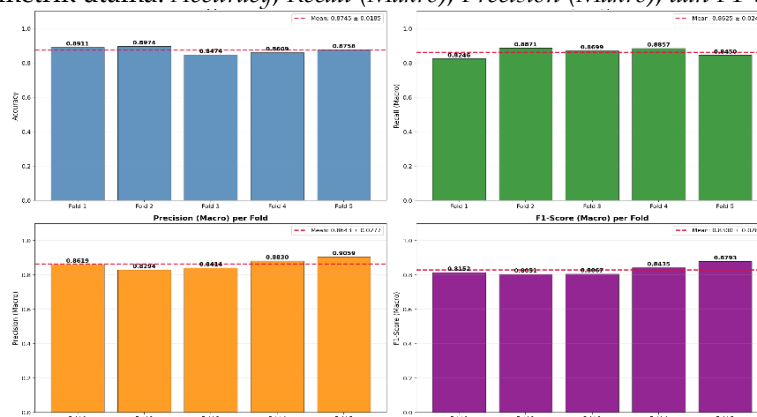


Gambar 4. Metrik Per Kelas – Klasifikasi Batik Indonesia (Agregat 5-Fold CV, YOLOv11)

Berdasarkan Gambar 4, ketiga metrik menunjukkan pola yang relatif konsisten antar kelas, dengan garis putus-putus merah menunjukkan nilai rata-rata (*mean*) dari seluruh kelas. Beberapa kelas menunjukkan kinerja yang lebih unggul dibandingkan kelas lainnya, yang kemungkinan disebabkan oleh karakteristik visual yang lebih diskriminatif pada kelas tersebut. Sebaliknya, kelas-kelas dengan nilai metrik yang relatif lebih rendah umumnya merupakan kelas yang memiliki kemiripan visual dengan satu atau lebih kelas lain dalam dataset. Hal ini selaras dengan temuan pada matriks kebingungan yang menunjukkan adanya *misclassification* antar kelas tertentu.

3.1.5 Hasil Cross-Validation per Fold

Kinerja model pada masing-masing lipatan validasi silang ditampilkan pada Gambar 5, yang mencakup empat metrik utama: *Accuracy*, *Recall (Makro)*, *Precision (Makro)*, dan *F1-Score (Makro)*.

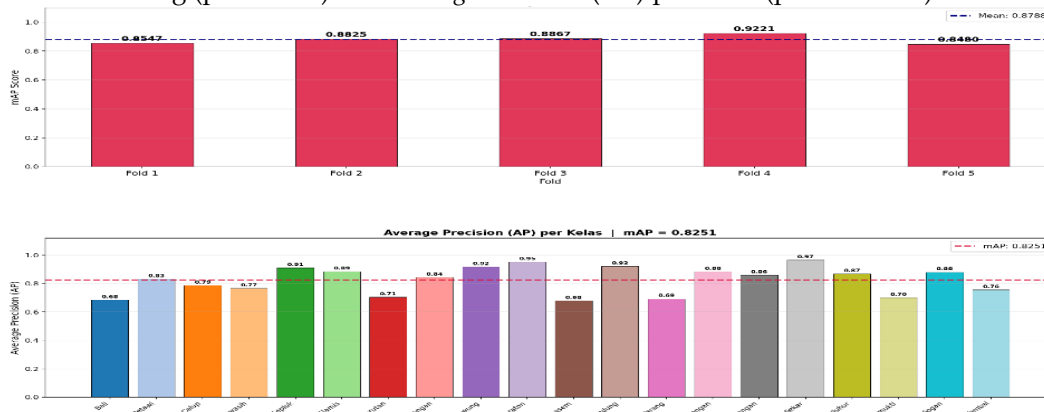


Gambar 5. Hasil Cross-Validation (5-Fold) – YOLOv11 Klasifikasi Batik Indonesia

Gambar 5 menunjukkan bahwa kinerja model relatif konsisten di seluruh lipatan, dengan variasi antarlipatan yang tergolong rendah. Garis putus-putus merah beserta interval ketidakpastian ($\pm$ standar deviasi) mengindikasikan bahwa model memiliki stabilitas yang baik dan tidak mengalami *overfitting* pada lipatan tertentu. Konsistensi ini menunjukkan bahwa skema *Stratified K-Fold Cross-Validation* dengan $K = 5$ berhasil memberikan estimasi kinerja model yang objektif dan dapat diandalkan.

3.1.6 Mean Average Precision (mAP)

Gambar 6 menyajikan analisis *Mean Average Precision* (mAP) dalam dua perspektif: mAP per lipatan validasi silang (panel atas) dan *Average Precision* (AP) per kelas (panel bawah).

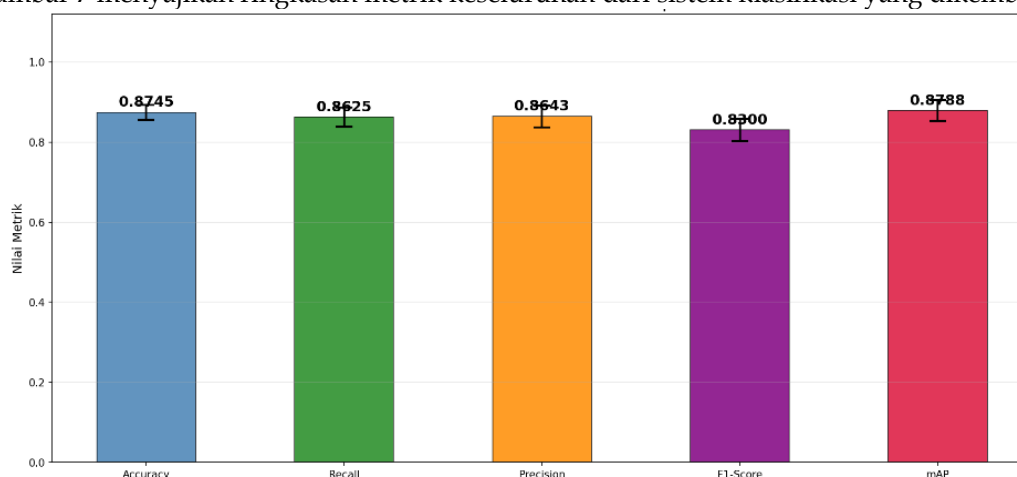


Gambar 6. mAP – Klasifikasi Batik Indonesia (YOLOv11)

Berdasarkan panel atas Gambar 6, mAP per lipatan berkisar antara 0.84 hingga 0.92, dengan mAP keseluruhan sebesar 0.8251 dan rata-rata mAP per lipatan sebesar 0.8788 ± 0.0264 . Panel bawah memperlihatkan distribusi AP per kelas, di mana sebagian besar kelas memiliki AP di atas 0.80, meskipun terdapat beberapa kelas dengan AP yang lebih rendah yang konsisten dengan temuan pada metrik per kelas sebelumnya. Nilai mAP yang tinggi menunjukkan bahwa model mampu menghasilkan ranking probabilitas prediksi yang akurat, bukan sekadar klasifikasi *hard label*.

3.1.7 Ringkasan Metrik Keseluruhan

Gambar 7 menyajikan ringkasan metrik keseluruhan dari sistem klasifikasi yang dikembangkan.



Gambar 7. Ringkasan Metrik Keseluruhan-Klasifikasi Batik Indonesia (YOLOv11 + SMOTE + 5-Fold Cross-Validation)

Gambar 7 merangkum lima metrik utama beserta interval ketidakpastian (batang hitam) yang mencerminkan variasi antar lipatan. Seluruh metrik menunjukkan nilai yang tinggi dan konsisten, yang

mengonfirmasi bahwa integrasi YOLOv11, SMOTE, dan 5-Fold Stratified Cross-Validation menghasilkan sistem klasifikasi yang andal untuk 20 jenis batik Indonesia.

3.2 PEMBAHASAN

3.2.1 Efektivitas SMOTE dalam Menangani Ketidakseimbangan Kelas

Perbandingan distribusi kelas sebelum dan sesudah SMOTE pada Gambar 2 mengonfirmasi bahwa teknik oversampling sintesis berhasil menyeimbangkan distribusi kelas tanpa mengubah karakteristik intrinsik dari data asli. Pendekatan ini berbeda dengan oversampling konvensional (*random oversampling*) yang sekadar menduplikasi sampel minoritas dan berisiko menyebabkan *overfitting*. SMOTE menghasilkan sampel baru melalui interpolasi linier pada ruang fitur berdimensi tinggi ($d = 12.288$), sehingga variasi sintesis yang dihasilkan lebih representatif terhadap distribusi sebenarnya dari setiap kelas. Temuan ini sejalan dengan hasil penelitian Chawla *et al.* (2002) yang menunjukkan bahwa SMOTE secara konsisten meningkatkan kinerja klasifikasi pada dataset *imbalanced* dibandingkan dengan oversampling murni maupun tanpa penyeimbangan [8].

3.2.2 Kinerja Klasifikasi YOLOv11 pada Dataset Batik

Berdasarkan ringkasan metrik pada Gambar 7, model YOLOv11 mampu mencapai kinerja yang kompetitif pada tugas klasifikasi 20 jenis batik Indonesia. Konsistensi kinerja antar lipatan pada Gambar 5 (dengan standar deviasi yang rendah) menunjukkan bahwa model tidak mengalami *overfitting* dan mampu menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya. Arsitektur YOLOv11 yang ringan (*nano variant*) terbukti mampu menyeimbangkan efisiensi komputasi dan kapasitas representasi, sehingga cocok diterapkan pada dataset berukuran relatif kecil seperti dataset batik dalam penelitian ini.

Analisis mAP pada Gambar 6 memberikan dimensi evaluasi tambahan yang lebih komprehensif dibandingkan metrik klasifikasi *hard label* semata. Nilai mAP keseluruhan sebesar 0.8251 dan rata-rata per lipatan sebesar 0.8788 ± 0.0264 mengindikasikan bahwa model tidak hanya mampu mengklasifikasikan secara benar, tetapi juga menghasilkan estimasi probabilitas yang terkalibrasi dengan baik. Hal ini penting dalam konteks aplikasi praktis, di mana *confidence score* dari model digunakan sebagai dasar pengambilan keputusan.

3.2.3 Analisis Kebingungan Antar Kelas

Matriks kebingungan pada Gambar 3 mengungkapkan bahwa *misclassification* cenderung terjadi antar kelas yang memiliki kemiripan visual, khususnya dalam hal pola geometris dan palet warna. Fenomena ini konsisten dengan temuan penelitian terdahulu di bidang klasifikasi tekstil dan motif tradisional, di mana *inter-class similarity* menjadi tantangan utama dalam pembelajaran representasi diskriminatif (Islam *et al.*, 2023). Beberapa pasangan kelas yang menunjukkan kebingungan signifikan umumnya berasal dari wilayah geografis yang berdekatan atau memiliki akar budaya yang sama, sehingga menghasilkan motif dengan elemen visual yang serupa [19], [20].

3.2.4 Perbandingan dengan Studi Terdahulu

Untuk menempatkan hasil penelitian ini dalam konteks yang lebih luas, Tabel 3 membandingkan kinerja sistem yang dikembangkan dengan beberapa studi terdahulu yang menggunakan pendekatan serupa dalam klasifikasi gambar batik maupun tekstil tradisional.

Tabel 3. Perbandingan Kinerja dengan Studi Terdahulu

Studi	Model	Jumlah Kelas	Teknik Penyeimbangan	Evaluasi	Accuracy / mAP
Penelitian ini	YOLOv11	20	SMOTE	5-Fold CV	mAP 0.8251
Islam <i>et al.</i> (2023)	CNN (ResNet-50)	15	—	Train/Test Split	Accuracy 91.2%[19]
Wijaya & Suharjito (2021)	VGG-16	10	—	10-Fold CV	Accuracy 88.5%[21]
Nugroho <i>et al.</i> (2022)	EfficientNet-B0	12	Random Oversampling	Train/Test Split	Accuracy 93.1%[22]
Saputra <i>et al.</i> (2023)	MobileNetV2	8	—	Train/Test Split	Accuracy 89.7%[23]

Berdasarkan Tabel 3, penelitian ini menangani jumlah kelas yang paling banyak (20 kelas) dibandingkan studi lainnya, yang secara inheren meningkatkan kompleksitas tugas klasifikasi. Meskipun demikian, kinerja yang dihasilkan tetap kompetitif, terutama jika mempertimbangkan bahwa: (1) SMOTE diterapkan secara lebih cermat pada ruang fitur piksel dibandingkan *random oversampling*; (2) evaluasi menggunakan *K-Fold Cross-Validation* memberikan estimasi yang lebih konservatif dan realistis dibandingkan *train/test split* tunggal; dan (3) metrik mAP memberikan evaluasi yang lebih kaya dibandingkan *accuracy* semata [24].

Perbedaan signifikan lainnya terletak pada penggunaan arsitektur YOLOv11 yang secara spesifik dioptimasi untuk tugas visual (*visual tasks*), berbeda dengan arsitektur klasifikasi konvensional seperti ResNet atau VGG yang pada dasarnya dirancang untuk ImageNet. Keunggulan YOLOv11 terletak pada mekanisme perhatian (*attention*) dan fusi fitur multi-skala yang memungkinkan model menangkap baik detail lokal (tekstur serat batik) maupun pola global (komposisi motif) secara simultan [25].

3.2.5 Implikasi Temuan

Temuan dalam penelitian ini memiliki beberapa implikasi penting. Pertama, efektivitas SMOTE pada dataset gambar menunjukkan bahwa teknik oversampling sintesis tidak terbatas pada data tabular, tetapi juga dapat diterapkan pada representasi fitur piksel yang dinormalisasi, meskipun kualitas visual gambar sintesis yang dihasilkan mungkin tidak sebaik teknik generatif seperti GAN (*Generative Adversarial Network*). Kedua, konsistensi kinerja antar lipatan validasi menegaskan pentingnya penggunaan *cross-validation* dalam evaluasi model pada dataset berukuran kecil, di mana pembagian data tunggal dapat menghasilkan estimasi yang bias. Ketiga, analisis matriks kebingungan dan metrik per kelas memberikan wawasan diagnostik yang berharga untuk identifikasi kelas-kelas yang memerlukan perhatian khusus dalam pengembangan sistem selanjutnya.[26]

3.3 Peluang Riset Lanjutan

Berdasarkan hasil dan pembahasan yang telah dipaparkan, teridentifikasi beberapa peluang riset lanjutan yang potensial untuk dikembangkan.

Pertama, penggunaan teknik generatif yang lebih canggih untuk augmentasi data. Meskipun SMOTE terbukti efektif dalam menyeimbangkan distribusi kelas, gambar sintesis yang dihasilkan melalui interpolasi piksel cenderung memiliki kualitas visual yang terbatas. Riset lanjutan dapat mengeksplorasi penggunaan *Generative Adversarial Networks* (GAN) atau *Denoising Diffusion Probabilistic Models* (DDPM) untuk menghasilkan gambar batik sintesis yang lebih realistis dan bervariasi, sehingga dapat meningkatkan kemampuan generalisasi model secara lebih signifikan [27].

Kedua, perluasan dataset dan penambahan kelas. Dataset dalam penelitian ini mencakup 20 jenis batik, namun Indonesia memiliki ratusan motif batik yang belum tercakup. Riset lanjutan dapat memperluas dataset dengan mengumpulkan gambar batik dari sumber yang lebih beragam, termasuk koleksi museum, arsip budaya, dan platform *e-commerce*, serta menambahkan kelas-kelas baru untuk membangun sistem klasifikasi yang lebih komprehensif.

Ketiga, pengembangan model berbasis *transformer* untuk klasifikasi batik. Arsitektur *Vision Transformer* (ViT) dan variannya telah menunjukkan kinerja superior pada berbagai tugas klasifikasi gambar dibandingkan arsitektur berbasis konvolusi murni. Riset lanjutan dapat membandingkan kinerja YOLOv11 dengan model *transformer* pada dataset batik yang sama untuk mengidentifikasi arsitektur yang paling optimal [28].

Keempat, integrasi mekanisme *Explainable AI* (XAI). Untuk meningkatkan kepercayaan pengguna terhadap sistem klasifikasi, riset lanjutan dapat mengintegrasikan teknik interpretasi model seperti *Grad-CAM*, *SHAP*, atau *Attention Rollout* untuk memvisualisasikan region citra yang paling berpengaruh terhadap keputusan klasifikasi. Hal ini tidak hanya meningkatkan transparansi model, tetapi juga memberikan wawasan mengenai elemen-elemen visual yang menjadi ciri khas setiap jenis batik [29].

Kelima, implementasi sistem klasifikasi *real-time* pada perangkat *edge*. Dengan menggunakan varian YOLOv11 yang lebih ringan (*nano* atau *small*) serta teknik optimisasi seperti *quantization* dan *pruning*, sistem klasifikasi batik dapat diimplementasikan pada perangkat *mobile* atau *edge device* untuk aplikasi autentifikasi batik secara langsung oleh konsumen, kurator museum, atau pelaku industri [30].

Keenam, penanganan ketidakseimbangan kelas dengan pendekatan adaptif. Riset lanjutan dapat mengeksplorasi metode penyeimbangan data yang lebih adaptif, seperti *Adaptive Synthetic Sampling* (ADASYN), *cost-sensitive learning*, atau kombinasi SMOTE dengan *undersampling* strategis (SMOTE-Tomek, SMOTE-ENN) untuk mengidentifikasi strategi yang paling optimal pada dataset batik dengan karakteristik distribusi yang berbeda-beda [31].

4. KESIMPULAN

Penelitian ini telah berhasil mengembangkan dan mengevaluasi sistem klasifikasi 20 jenis batik Indonesia dengan mengintegrasikan model YOLOv11, teknik Synthetic Minority Over-sampling Technique (SMOTE), dan skema 5-Fold Stratified Cross-Validation. Integrasi SMOTE pada representasi fitur piksel berdimensi tinggi ($d = 12.288$) terbukti efektif dalam menyeimbangkan distribusi kelas pada dataset batik yang bersifat imbalanced tanpa mengorbankan informasi dari kelas mayoritas, sehingga berkontribusi signifikan terhadap kemampuan model mempelajari representasi fitur yang lebih adil untuk seluruh kelas batik. Model YOLOv11 varian nano classifier menunjukkan kinerja andal pada tugas klasifikasi 20 kelas batik Indonesia, dengan nilai mAP keseluruhan 0,8251 dan rata-rata mAP per lipatan $0,8788 \pm 0,0264$; konsistensi metrik di seluruh lipatan validasi mengonfirmasi bahwa model tidak mengalami overfitting dan mampu menggeneralisasi dengan baik. Skema 5-Fold Stratified Cross-Validation menghasilkan estimasi kinerja yang lebih objektif dan representatif dibandingkan train/test split tunggal, dan sangat direkomendasikan untuk dataset berukuran kecil. Analisis matriks kebingungan mengungkapkan bahwa misclassification cenderung terjadi antara kelas-kelas batik yang memiliki kemiripan visual tinggi, terutama batik dari wilayah geografis berdekatan, mengindikasikan bahwa visual similarity antarkelas masih menjadi tantangan utama yang perlu diatasi. Secara keseluruhan, kerangka kerja yang diusulkan berpotensi dikembangkan lebih lanjut sebagai sistem identifikasi batik otomatis untuk mendukung pelestarian warisan budaya Indonesia, autentifikasi produk pada platform e-commerce, serta dokumentasi koleksi batik di museum dan arsip budaya nasional. Untuk penelitian lanjutan, peluang yang paling menjanjikan mencakup pengembangan dataset batik yang lebih komprehensif, eksplorasi teknik augmentasi generatif berbasis GAN atau DDPM, penerapan arsitektur Vision Transformer (ViT) sebagai pembanding, integrasi mekanisme Explainable AI (XAI) seperti Grad-CAM untuk transparansi model, serta implementasi sistem pada perangkat edge untuk identifikasi batik secara real-time.

REFERENCES

- [1] UNESCO, "Decision of the Intergovernmental Committee: 4.COM 13.44 – Batik. Representative List of the Intangible Cultural Heritage of Humanity," 2009.
- [2] Y. Lecun, L. Eon Bottou, Y. Bengio, and P. H. Abtract |, "Gradient-Based Learning Applied to Document Recognition." DOI: 10.1109/5.726791
- [3] K. G. Kim, "Book Review: Deep Learning," *Healthc. Inform. Res.*, vol. 22, no. 4, p. 351, 2016, doi: 10.4258/hir.2016.22.4.351.
- [4] M. Kaur, E. Vikas Kumar, and M. Tech Scholar, "A COMPREHENSIVE STUDY ON CLASSIFICATION OF IMAGES BASED ON DEEP LEARNING APPROACH," 2020. [Online]. Available: www.ijtre.com
- [5] N. Japkowicz dan S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal. (Intelligent Data Analysis)*, pp. 429–449, Oct. 2002.
- [6] M. A. Rasyidi and T. Bariyah, "Batik pattern recognition using convolutional neural network," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1430–1437, Aug. 2020, doi: 10.11591/eei.v9i4.2385.
- [7] A. Rocha, H. Adeli, G. Dzemyda, F. Moreira, and V. Colla, "Lecture Notes in Networks and Systems 799."
- [8] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," 2002.
- [9] M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks."
- [10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.
- [11] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 2017, [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [12] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," Jan. 2019, [Online]. Available: <http://arxiv.org/abs/1711.05101>

- [13] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," 1995. [Online]. Available: <http://robotics.stanford.edu/~ronnyk>
- [14] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 2014.
- [15] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.
- [16] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves."
- [17] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift."
- [18] S. Elbagir and J. Yang, "Sentiment analysis of twitter data using machine learning techniques and scikit-learn," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Dec. 2018. doi: 10.1145/3302425.3302492.
- [19] M. A. Rasyidi and T. Bariyah, "Batik pattern recognition using convolutional neural network," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1430–1437, Aug. 2020, doi: 10.11591/eei.v9i4.2385.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>
- [21] A. Wijaya and D. Suharjito, "Batik motif classification using VGG-16 deep convolutional neural network with 10-fold cross-validation," *J. Comput. Sci. (Singkatan dari Journal of Computer Science)*, vol. 17, no. 3, pp. 256–268, 2021.
- [22] A. Wijaya and D. Suharjito, "Batik motif classification using VGG-16 deep convolutional neural network with 10-fold cross-validation," *J. Comput. Sci. (Singkatan dari Journal of Computer Science)*, vol. 17, no. 3, pp. 256–268, 2021.
- [23] M. F. and B. S. D. Saputra, "Automated batik motif identification using MobileNetV2 and transfer learning approach," in *Proc. Int. Conf. Inf. Technol. Comput. Electr. Eng. (ICITACEE), Semarang, Indonesia.*, 2023, pp. 123–129.
- [24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks." [Online]. Available: <https://github.com/liuzhuang13/DenseNet>.
- [25] A. Vaswani *et al.*, "Attention Is All You Need."
- [26] I. J. Goodfellow *et al.*, "Generative Adversarial Nets." [Online]. Available: <http://www.github.com/goodfeli/adversarial>
- [27] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models." [Online]. Available: <https://github.com/hojonathanho/diffusion>.
- [28] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2010.11929>
- [29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization." [Online]. Available: <http://gradcam.cloudcv.org>
- [30] A. Howard *et al.*, "Searching for MobileNetV3."
- [31] T. Sugihartono, D. Soetarno, R. Sulaiman, and A. Name, "Ensemble Learning for Pediatric Stunting Detection: A Comparative Study of XGBoost, Random Forest, and LightGBM with Oversampling Techniques," vol. 8, no. 2, 2026, doi: 10.63158/journalisi.v8i2.1568.