

Klasifikasi Motif Batik Nusantara Menggunakan Transfer Learning EfficientNetB0

Zainul Mutawakkil[1], Mohammad Alwan Abbas Abdillah[2], Helwa Nuftiana Fadoil Putri[3*], Hasna Kamalia Hakim[4], Fikri Amarulloh Tyto Pratama[5], Latif Khanan Al-Mansuri[6]

Program Studi Informatika, Fakultas Matematika dan Ilmu Komputer, Universitas Nahdlatul Ulama Al-Ghazali Cilacap, Cilacap, Indonesia

Email: elind.zainul@gmail.com[1], alwanmuhamad580@gmail.com[2], helwaputri55@gmail.com[3], hasnakamalia761@gmail.com[4], fikripratamaclp2019@gmail.com[5], latifkhanan@gmail.com[6]

Abstrak

Penelitian ini mengembangkan sistem klasifikasi citra untuk mengenali 20 motif batik Nusantara dengan transfer learning EfficientNetB0. Masalah penelitian berangkat dari sulitnya membedakan motif batik yang memiliki struktur visual mirip, variasi warna, perubahan pencahayaan, kualitas kamera, dan sudut pengambilan gambar. Dataset lokal yang digunakan berasal dari Kaggle Batik Nusantara dan mencakup 800 citra: 560 citra pelatihan, 80 citra validasi, dan 160 citra test independen. Validasi digunakan untuk pemilihan bobot model, sedangkan kemampuan generalisasi akhir dinilai pada test independen. Metode penelitian meliputi persiapan dataset, preprocessing citra 224 x 224 piksel, augmentasi numerik, pelatihan dua tahap, evaluasi kuantitatif, analisis kesalahan, dan integrasi model ke aplikasi web Flask. Model dilatih dengan ekstraksi fitur dan fine-tuning, menggunakan categorical crossentropy dengan label smoothing 0,1, optimizer Adam, ModelCheckpoint, EarlyStopping, dan ReduceLROnPlateau. Evaluasi menghasilkan akurasi validasi 81,25%, sedangkan evaluasi pada test independen menghasilkan akurasi 78,75%, macro precision 0,8170, macro recall 0,7875, macro F1-score 0,7843, dan test loss 1,1609. Hasil ini menunjukkan bahwa EfficientNetB0 dapat menjadi *baseline transfer learning* yang layak untuk klasifikasi motif batik pada dataset kecil, dengan penurunan performa yang relatif terbatas saat diuji pada data yang tidak dipakai untuk *tuning*.

Kata Kunci: Aplikasi Web; EfficientNetB0; Klasifikasi Batik; Pembelajaran Transfer; Visi Komputer

ABSTRACT

This study develops an image classification system to recognize 20 Nusantara batik motifs using EfficientNetB0 transfer learning. The research problem arises from the difficulty of distinguishing visually similar batik motifs under variations in color, lighting, camera quality, and shooting angle. The local dataset was derived from the Kaggle Batik Nusantara dataset and contains 800 images: 560 training images, 80 validation images, and 160 independent test images. Validation was used for model selection, while final generalization was assessed on the independent test set. The method covers dataset preparation, 224 x 224 image preprocessing, numerical augmentation, two-stage training, quantitative evaluation, error analysis, and Flask web integration. The model was trained through feature extraction and fine-tuning using categorical crossentropy with 0.1 label smoothing, Adam optimizer, ModelCheckpoint, EarlyStopping, and ReduceLROnPlateau. The experiment produced 81.25% validation accuracy, while the independent test set produced 78.75% accuracy, 0.8170 macro precision, 0.7875 macro recall, 0.7843 macro F1-score, and 1.1609 test loss. These findings indicate that EfficientNetB0 is a feasible transfer learning baseline for batik motif classification on a small dataset, with only a modest drop when evaluated on unseen data.

Keywords: *Web application; EfficientNetB0; Batik classification; Transfer learning; Vision computing*

1. PENDAHULUAN

Batik adalah warisan budaya dengan keanekaragaman motif, warna, dan arti simbolis. Setiap wilayah dapat memiliki ciri visual yang berbeda, seperti bentuk geometris, tumbuhan, hewan, hiasan rumah tradisional, dan pola yang berulang. Keberagaman itu menjadikan pengenalan motif batik sebagai tugas yang memerlukan pengalaman visual. Dalam konteks digital, tuntutan untuk identifikasi otomatis semakin krusial karena kumpulan citra batik dapat dimanfaatkan untuk pendidikan, pelestarian budaya, pencarian visual, dan pengembangan aplikasi yang berbasis kecerdasan buatan. Kajian mengenai produksi batik warna alam, dinamika batik Indonesia-Malaysia, dan posisi batik dalam industri kreatif menunjukkan bahwa digitalisasi batik tidak hanya terkait klasifikasi citra, tetapi juga berkaitan dengan pelestarian praktik produksi, nilai budaya, dan diplomasi budaya kontemporer (Handayani et al., 2018; Hajura, 2022; Shaharuddin et al., 2021).

Masalah klasifikasi motif batik muncul karena beberapa motif memiliki struktur garis, ornamen, dan pengulangan pola yang mirip. Perubahan pencahayaan, ukuran objek, lipatan kain, mutu kamera, latar belakang, dan perbedaan warna dapat menggeser fitur visual yang dibaca model. Klasifikasi manual juga bergantung pada pengalaman pengamat, sehingga hasilnya dapat berbeda antar pengguna. Kondisi ini menuntut sistem klasifikasi citra yang mampu mengekstraksi fitur spasial secara konsisten, tetapi tetap perlu diuji secara ketat karena motif batik tidak hanya berbeda secara visual, melainkan juga terkait dengan konteks budaya. Penelitian generatif mengenai motif batik menunjukkan bahwa model *deep learning* dapat mempelajari representasi visual batik, tetapi data yang terbatas dapat membuat representasi tersebut kurang stabil pada motif yang mirip (Octadion et al., 2023).

Convolutional Neural Network banyak diaplikasikan dalam klasifikasi gambar karena dapat belajar fitur lokal melalui proses konvolusi. Akan tetapi, CNN yang dirancang dari nol membutuhkan data dalam jumlah besar agar tidak rentan terhadap *overfitting*. Kondisi ini menjadi persoalan pada dataset batik yang sering berukuran kecil. Strategi pembelajaran transfer menjadi jawaban yang tepat karena model yang telah dilatih sebelumnya telah memahami ciri-ciri umum dari dataset yang besar, kemudian disesuaikan untuk domain baru melalui ekstraksi fitur dan penyempurnaan. Penelitian *transfer learning* menunjukkan bahwa pemindahan fitur dari model yang telah dilatih sebelumnya dapat meningkatkan efisiensi pelatihan dan performa model pada domain target dengan data yang terbatas (Jiang et al., 2022; Plested et al., 2025).

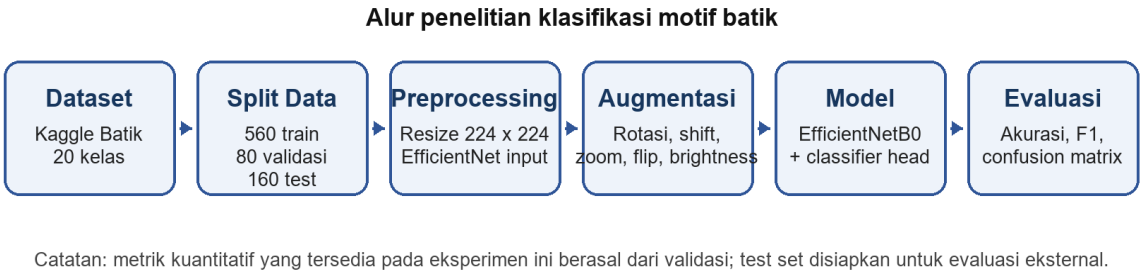
Penelitian sebelumnya telah menunjukkan bahwa CNN, *transfer learning*, augmentasi data, dan model visi modern dapat meningkatkan klasifikasi citra. Namun, masih terdapat celah pada penerapan klasifikasi motif batik berskala kecil yaitu banyak studi menekankan nilai akurasi akhir tanpa menjelaskan pembagian data secara rinci, tanpa *test set* independen, tanpa analisis kesalahan per kelas yang memadai, dan tanpa justifikasi pemilihan *backbone*. Celah ini penting karena dataset batik sering berukuran kecil dan memiliki variasi intra-kelas yang tinggi. EfficientNetB0 dipilih sebagai *baseline* karena menggunakan *compound scaling* yang menyeimbangkan kedalaman, lebar, dan resolusi jaringan (Tan & Le, 2019). Pemilihan ini bukan klaim bahwa EfficientNetB0 paling unggul dibandingkan ResNet, MobileNet, atau ConvNeXt, melainkan keputusan metodologis yang sesuai untuk eksperimen awal dengan sumber daya komputasi terbatas. ConvNeXt dan kajian model visi modern tetap relevan sebagai pembanding lanjutan karena menunjukkan bahwa desain CNN kontemporer dapat bersaing dengan arsitektur transformer (Han et al., 2023; Khan et al., 2022; Liu et al., 2022).

Studi ini bertujuan untuk menjawab tiga pertanyaan penelitian. Pertama adalah bagaimana merancang *pipeline* klasifikasi 20 motif batik Nusantara berbasis transfer learning EfficientNetB0 pada dataset kecil. Kedua adalah bagaimana kinerja model berdasarkan akurasi, *precision, recall, F1-score, loss*, dan *confusion matrix*. Ketiga adalah kelas motif apa yang paling mudah dan paling sulit diklasifikasikan, serta faktor apa yang diduga menyebabkan kesalahan tersebut. Kontribusi penelitian ini terletak pada dokumentasi *pipeline*, pelaporan pembagian data, analisis performa per kelas, dan identifikasi batas validitas model pada dataset batik yang terbatas.

2. METODE

2.1 Desain Penelitian

Studi ini menerapkan pendekatan eksperimen berbasis komputer. Objek penelitian adalah citra motif batik Nusantara yang dibagi ke dalam 20 kategori. Model yang digunakan adalah EfficientNetB0 pretrained ImageNet. Proses penelitian terdiri atas pengumpulan data, pembagian data, *preprocessing*, augmentasi, pembangunan model, pelatihan dua tahap, evaluasi validasi, evaluasi test independen, analisis kesalahan, dan integrasi aplikasi web. Alur tersebut ditunjukkan pada Gambar 1. Validation set digunakan untuk memilih bobot terbaik, sedangkan test set independen digunakan untuk menilai generalisasi model.



Gambar 1. Diagram Alur Penelitian Klasifikasi Motif Batik Nusantara

Sistem penelitian dirancang dengan cara yang modular. File konfigurasi menyimpan *hyperparameter*, data loader mengelola dataset dan generator gambar, file model membangun arsitektur EfficientNetB0, *file train* melaksanakan pelatihan, *file evaluate* menghasilkan metrik evaluasi, sementara *file predict* digunakan oleh CLI dan aplikasi Flask. Pemisahan modul ini mendukung pengujian, perbaikan, dan pengembangan yang lebih lanjut.

2.2 Dataset dan Preprocessing

Dataset yang digunakan berasal dari Kaggle dengan nama Batik Nusantara atau Batik Indonesia Dataset. Struktur data lokal terdiri atas *folder train* dan *test*. *Folder train* memuat 640 citra dari 20 kelas, dengan 32 citra per kelas. Pada proses pelatihan, *folder train* dibagi menjadi 560 *citra training* dan 80 citra validasi, sehingga setiap kelas memiliki 28 *citra training* dan 4 citra validasi. Folder test memuat 160 citra independen atau 8 citra per kelas. Set test ini tidak digunakan dalam ModelCheckpoint, EarlyStopping, atau pemilihan bobot model, tetapi digunakan pada evaluasi akhir untuk menilai generalisasi. Setiap citra diubah menjadi 224 x 224 piksel dengan tiga saluran RGB agar sesuai dengan input EfficientNetB0. Pra-pemrosesan menggunakan fungsi preprocess_input dari EfficientNet agar distribusi nilai piksel selaras dengan bobot ImageNet.

Untuk menurunkan risiko *overfitting, data training* diberi augmentasi numerik. Parameter yang digunakan meliputi rotation_range 30 derajat, width_shift_range 0,25, height_shift_range 0,25, shear_range 0,20, zoom_range 0,20, horizontal_flip, vertical_flip, brightness_range 0,8 sampai 1,2, channel_shift_range 10, dan fill_mode nearest. Data validasi tidak diberi augmentasi, tetapi tetap melalui preprocessing yang sama. Strategi ini dipilih karena setiap kelas hanya memiliki 28 citra training setelah split, sehingga model berisiko menghafal sampel pelatihan. Literatur augmentasi data menunjukkan bahwa variasi buatan dapat membantu generalisasi model pada tugas klasifikasi citra, terutama ketika jumlah data asli terbatas (Kumar et al., 2023; Sapkota et al., 2025; Yang et al., 2022).

2.3 Arsitektur Model

Model dikembangkan dengan EfficientNetB0 sebagai ekstraktor fitur. *Classifier* asli EfficientNetB0 tidak digunakan. Bagian akhir model diganti dengan classifier head yang terdiri atas GlobalAveragePooling2D, Dense 256 unit, BatchNormalization, aktivasi ReLU, Dropout 0,5, dan Dense softmax sesuai jumlah kelas. Pada tahap pertama, seluruh lapisan EfficientNetB0 dibekukan sehingga hanya classifier head yang dilatih. Pada tahap kedua*, fine-tuning* dilakukan dengan membuka lapisan mulai indeks ke-80. Keputusan ini bersifat eksploratif dan bertujuan menjaga fitur awal yang umum tetap stabil, sambil memberi ruang adaptasi pada blok akhir yang lebih spesifik terhadap tekstur. Pemilihan layer ke-80 belum merupakan hasil optimasi arsitektural, sehingga perlu diuji ulang melalui studi ablasi pada penelitian berikutnya.

Loss function yang diterapkan adalah *categorical crossentropy* dengan label *smoothing* sebesar 0,1. Label *smoothing* dipilih untuk menurunkan tingkat kepercayaan berlebih pada prediksi model, terutama disebabkan karena jumlah data per kelas yang terbatas. *Optimizer* yang dipakai adalah Adam. Proses pelatihan juga memanfaatkan ModelCheckpoint, EarlyStopping, dan ReduceLROnPlateau. Penelitian terkini menunjukkan bahwa label smoothing dapat mendukung kalibrasi prediksi dan generalisasi model dalam situasi data yang terbatas, meskipun pengaturannya masih harus disesuaikan dengan sifat dataset (Wang et al., 2024; Wei et al., 2022).

Tabel 1. Parameter Utama Penelitian

Parameter	Nilai
Sumber dataset	Kaggle Batik Nusantara / Batik Indonesia Dataset
Jumlah kelas	20 motif batik
Total citra lokal	800 citra
Data training	560 citra (28 citra per kelas)
Data validasi	80 citra (4 citra per kelas)
Data test independen	160 citra (8 citra per kelas, digunakan untuk evaluasi akhir)
Ukuran input	224 x 224 x 3
Backbone	EfficientNetB0 pretrained ImageNet
Classifier head	GAP, Dense 256, BatchNormalization, ReLU, Dropout 0,5, Softmax
Batch size	32
Epoch maksimum	80
Epoch berjalan	30
Fine-tuning	Unfreeze mulai layer ke-80; perlu studi ablasi lanjutan

Tabel 1 merangkum parameter inti penelitian, mulai dari sumber dataset, jumlah kelas, pembagian data, hingga ukuran input citra yang digunakan model. Ringkasan ini menegaskan bahwa desain eksperimen disusun secara seimbang pada 20 kelas, dengan pemisahan yang jelas

antara 560 citra training, 80 citra validasi, dan 160 citra test independen. Dengan skema ini, hasil validasi dapat dipakai untuk tuning model, sedangkan hasil test independen dapat dibaca sebagai ukuran generalisasi yang lebih layak untuk pelaporan ilmiah.

2.4 Evaluasi Model

Evaluasi kuantitatif dilakukan pada dua tahap. Tahap pertama memakai validation set yang berjumlah 80 citra untuk pemilihan bobot terbaik melalui ModelCheckpoint dan EarlyStopping. Tahap kedua memakai test set independen yang berjumlah 160 citra untuk menilai kemampuan generalisasi model. Metrik yang digunakan meliputi akurasi, *precision, recall, F1-score, loss*, dan *confusion matrix*. Akurasi menilai proporsi prediksi benar terhadap seluruh data. *Precision* menggambarkan ketepatan prediksi ketika model memilih sebuah kelas. *Recall* menggambarkan kemampuan model menemukan seluruh sampel dari kelas tertentu. F1-score menggabungkan *precision* dan *recall. Confusion matrix* digunakan untuk membaca pola kesalahan antar kelas. Dalam artikel ini, hasil validasi tetap dilaporkan sebagai bagian dari proses *tuning*, tetapi hasil utama generalisasi diambil dari test independen.

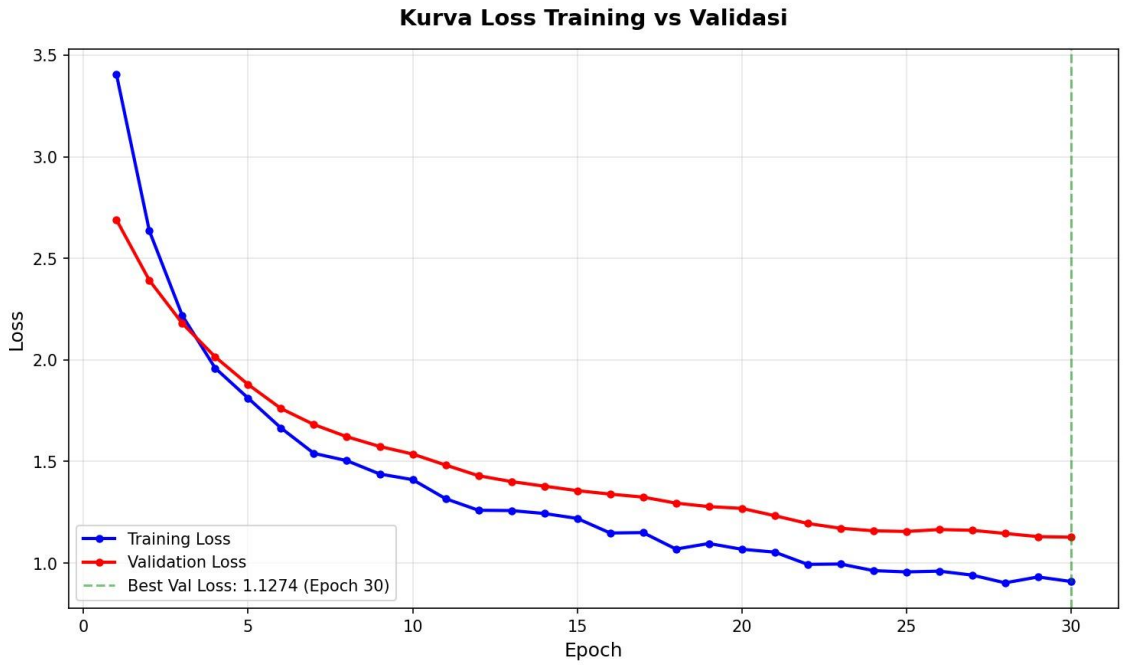
Implementasi aplikasi web ditempatkan sebagai keluaran terapan, bukan bukti utama performa model. Bukti ilmiah utama tetap berasal dari metrik validasi, tabel evaluasi, dan analisis kesalahan per kelas. Aplikasi Flask digunakan untuk menunjukkan bahwa model dapat dihubungkan ke antarmuka pengguna melalui input unggahan gambar dan kamera.

3. HASIL DAN PEMBAHASAN

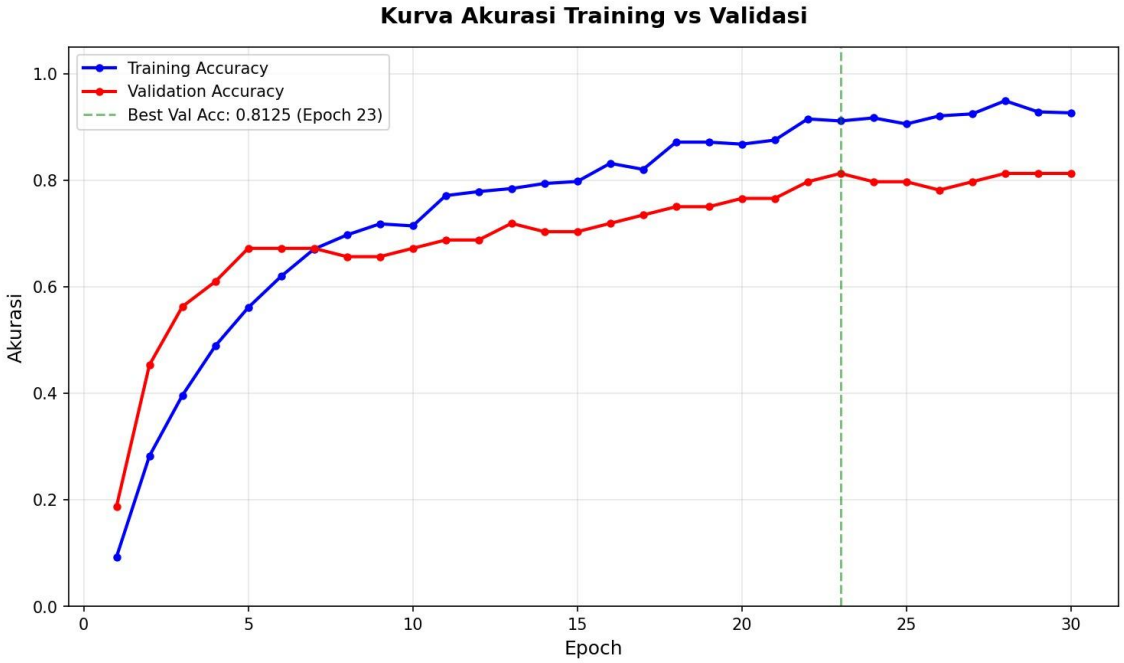
3.1 Hasil Pelatihan

Pelatihan model berlangsung selama 30 epoch dari batas maksimum 80 epoch. Bobot terbaik berdasarkan *validation accuracy* tercatat pada epoch ke-23. Akurasi akhir training mencapai 92,61%, sedangkan akurasi validasi berada pada 81,25%. Selisih sekitar 11,36 poin persentase menunjukkan *overfitting* ringan hingga sedang. Pola ini wajar pada dataset kecil dengan 20 kelas, tetapi tetap menjadi sinyal bahwa generalisasi model belum kuat.

Gambar 2 menunjukkan kurva *loss training* dan validasi. *Training loss* turun dari 3,4058 pada epoch pertama menjadi 0,9096 pada epoch ke-30. *Validation loss* juga turun dari 2,6899 menjadi 1,1274. Penurunan validation loss menunjukkan bahwa model masih belajar representasi yang berguna. Namun, jarak antara *training loss* dan *validation loss* tetap terlihat, sehingga model lebih baik menjelaskan data training daripada data validasi.



Gambar 2. Kurva Loss Training dan Validasi



Gambar 3. Kurva Akurasi Training dan Validasi

Gambar 3 memperlihatkan peningkatan akurasi *training* dan validasi. Akurasi *training* naik tajam pada awal pelatihan dan mencapai nilai tertinggi 94,89% pada epoch ke-28. Akurasi validasi mencapai nilai tertinggi 81,25% pada epoch ke-23, lalu cenderung mendatar sampai akhir pelatihan. Setelah epoch ke-23, peningkatan akurasi *training* tidak lagi diikuti peningkatan validasi. Titik ini menunjukkan awal saturasi generalisasi dan menjadi alasan mengapa mekanisme checkpoint memilih bobot terbaik berdasarkan *validation accuracy*.

3.2 Hasil Evaluasi Kuantitatif

Tabel 2. Ringkasan Hasil Evaluasi Model pada Validasi dan Test Independen

Metrik	Nilai
Validation loss	1,1800
Validation accuracy	0,8125 atau 81,25%
Test loss	1,1609
Test accuracy	0,7875 atau 78,75%
Macro precision test	0,8170
Macro recall test	0,7875
Macro F1-score test	0,7843
Weighted precision test	0,8170
Weighted recall test	0,7875
Weighted F1-score test	0,7843
Support test	160 citra

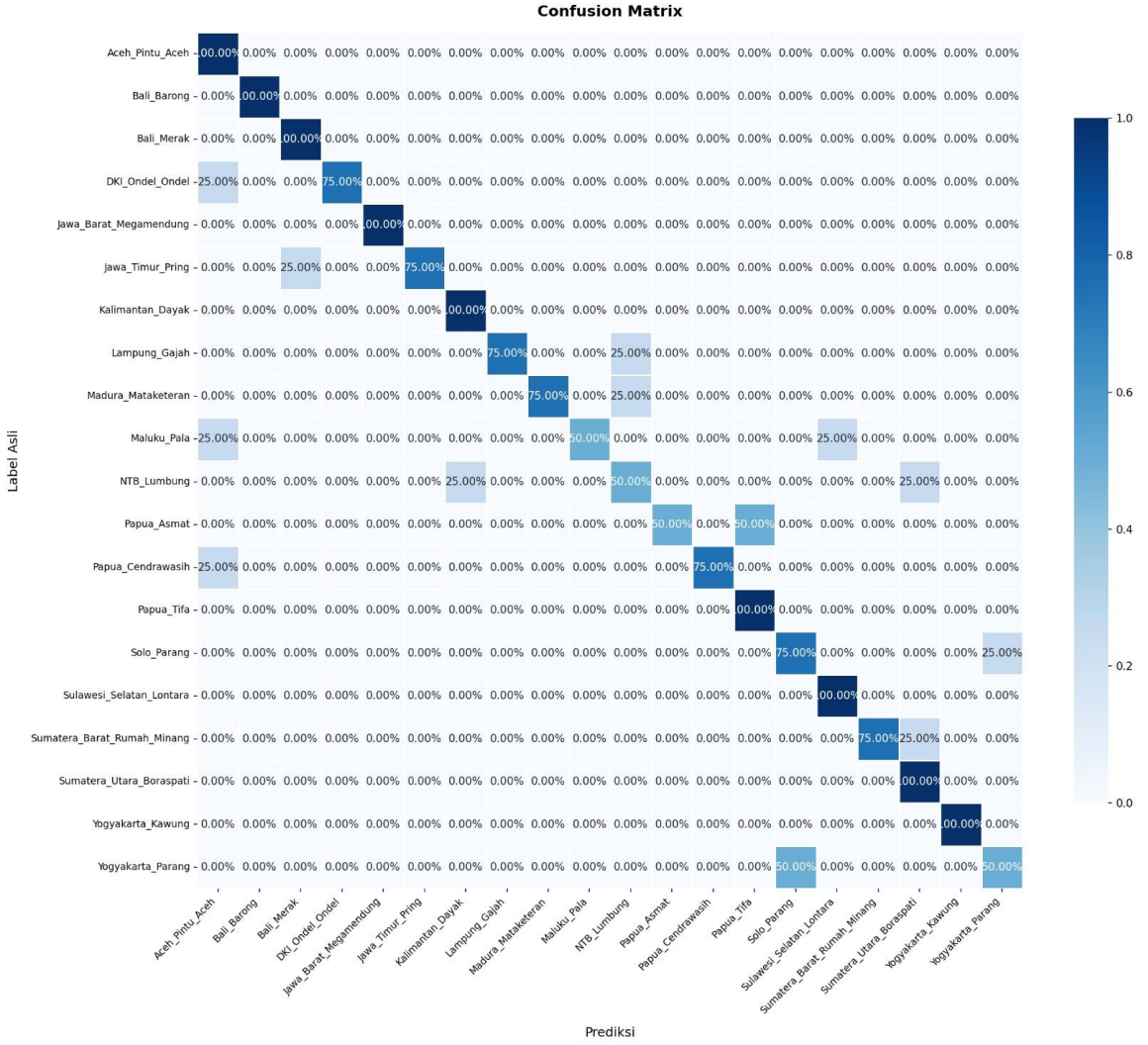
Tabel 2 merangkum hasil evaluasi kuantitatif pada *validation set* dan *test set* independen. Pada tahap validasi, model menghasilkan validation loss 1,1800 dan *validation accuracy* 81,25%. Pada test independen, akurasi turun menjadi 78,75% dengan *test loss* 1,1609, *macro precision* 0,8170, *macro recall* 0,7875, dan macro *F1-score* 0,7843. Penurunan sebesar 2,5 poin persentase dari validasi ke *test* menunjukkan bahwa performa model masih relatif stabil pada data yang tidak digunakan untuk *tuning*, meskipun beberapa kelas tetap menampilkan sensitivitas yang lebih rendah.

3.3 Analisis Per Kelas

Tabel 3. Kelas dengan Performa Tinggi dan Rendah pada Test Independen

Kategori	Kelas	*Precision*	*Recall*	*F1-score*
Tinggi	Jawa_Barat_Megamendung	1,0000	1,0000	1,0000
Tinggi	Maluku_Pala	1,0000	1,0000	1,0000
Tinggi	Yogyakarta_Kawung	1,0000	1,0000	1,0000
Rendah	Papua_Tifa	1,0000	0,3750	0,5455
Rendah	Aceh_Pintu_Aceh	0,6667	0,5000	0,5714
Rendah	Jawa_Timur_Pring	0,7143	0,6250	0,6667
Rendah	NTB_Lumbung	0,5385	0,8750	0,6667
Rendah	Yogyakarta_Parang	0,7143	0,6250	0,6667

Tabel 3 membandingkan kelas dengan performa tinggi dan rendah pada test independen. Kelas Jawa_Barat_Megamendung, Maluku_Pala, dan Yogyakarta_Kawung memperoleh *precision, recall*, dan *F1-score* sebesar 1,0000. Kalimantan_Dayak dan Lampung_Gajah juga menunjukkan performa kuat dengan F1-score di atas 0,93. Sebaliknya, Papua_Tifa, Aceh_Pintu_Aceh, Jawa_Timur_Pring, NTB_Lumbung, dan Yogyakarta_Parang memiliki F1-score yang lebih rendah. Pola ini menunjukkan bahwa kualitas generalisasi tidak seragam antar kelas. Beberapa kelas tetap stabil pada data baru, sedangkan kelas lain masih sensitif terhadap kemiripan motif, variasi sudut, atau variasi tekstur.



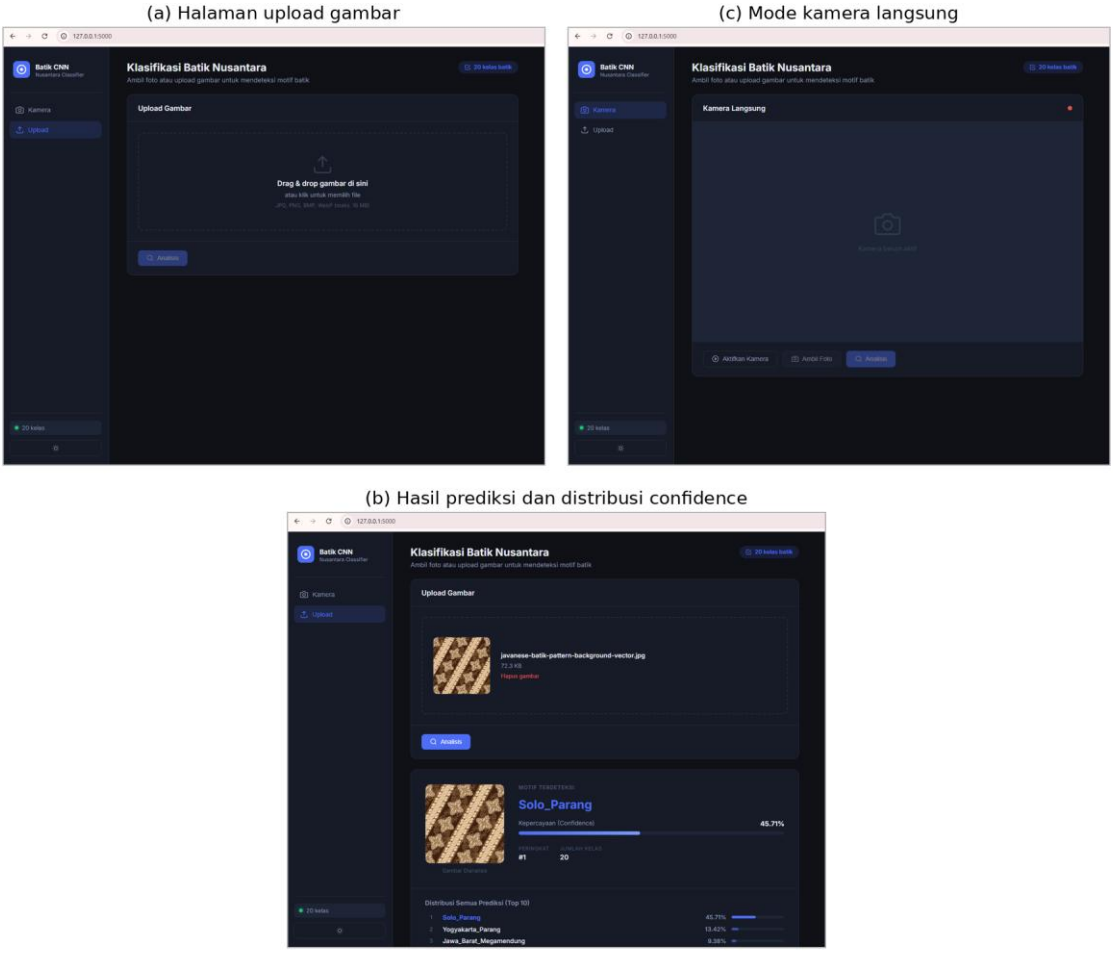
Gambar 4. Confusion Matrix Klasifikasi 20 Motif Batik pada Test Independen

Gambar 4 menunjukkan *confusion matrix* pada *test independen*. Diagonal utama tetap mendominasi, yang berarti sebagian besar kelas masih diprediksi dengan benar pada data yang tidak dipakai saat *tuning*. Namun, kesalahan yang lebih terlihat muncul pada kelas seperti Papua_Tifa, Aceh_Pintu_Aceh, NTB_Lumbung, Jawa_Timur_Pring, dan Yogyakarta_Parang. Secara lebih spesifik, Papua_Asmat diprediksi sebagai Bali_Barong sebesar 37,5%, sedangkan Solo_Parang dan Yogyakarta_Parang saling tertukar masing-masing sebesar 25,0%, yang menunjukkan bahwa motif dengan ornamen berulang atau arah garis diagonal yang serupa masih sulit dibedakan oleh model. Hasil ini menegaskan bahwa model tidak runtuh ketika berpindah dari validasi ke test, tetapi beberapa pasangan kelas masih sulit dibedakan karena memiliki pola ornamen dan tekstur yang serupa. Karena setiap kelas pada *test set* memiliki delapan citra, *confusion matrix* ini memberikan dasar evaluasi generalisasi yang lebih kuat dibandingkan confusion matrix validasi.

3.4 Implementasi Aplikasi Web

Model yang telah dilatih diintegrasikan ke dalam aplikasi web Flask sebagai prototipe penggunaan. Aplikasi menerima dua jenis input, yaitu unggahan gambar dan kamera langsung. Pada mode unggahan, pengguna memilih berkas JPG, PNG, BMP, atau WebP. Pada mode kamera, browser menangkap frame, mengubahnya menjadi base64, lalu mengirimkannya ke endpoint prediksi. Server memuat model sekali dan memakai cache agar permintaan berikutnya lebih cepat.

Hasil prediksi ditampilkan sebagai kelas utama, confidence, dan daftar sepuluh prediksi teratas. Fitur ini berguna untuk demonstrasi, tetapi tidak menggantikan evaluasi ilmiah. Dalam konteks penelitian, aplikasi web berperan sebagai bukti implementasi sistem, sedangkan validitas model tetap harus dinilai melalui data validasi, test independen, dan analisis kesalahan.



Gambar 5. Tampilan Antarmuka Aplikasi Web Klasifikasi Motif Batik

Gambar 5 menunjukkan antarmuka aplikasi web yang dipakai untuk menguji model secara langsung. Panel unggahan gambar, hasil prediksi, dan mode kamera memperlihatkan

bahwa pipeline inferensi telah dihubungkan ke sistem interaktif. Namun, tampilan aplikasi tidak digunakan sebagai dasar klaim akurasi model.

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa EfficientNetB0 dapat digunakan sebagai *baseline transfer learning* untuk klasifikasi 20 motif batik pada dataset kecil. Akurasi validasi 81,25% dan akurasi *test* independen 78,75% menunjukkan bahwa model mempertahankan performa yang relatif dekat saat berpindah dari data tuning ke data yang tidak terlihat sebelumnya. Hal ini penting karena banyak eksperimen pada dataset kecil hanya berhenti pada validasi. Namun, hasil tersebut belum cukup untuk menyatakan keunggulan efisiensi model atau kesiapan penggunaan umum karena belum ada pembanding empiris dengan ResNet, MobileNet, ConvNeXt, atau EfficientNetV2 pada dataset yang sama. Kontribusi utama penelitian ini terletak pada pipeline, pelaporan dua tahap evaluasi, dan analisis kesalahan per kelas.

Keterbatasan utama penelitian tetap berada pada ukuran dataset. Setiap kelas hanya memiliki 28 *citra training*, 4 citra validasi, dan 8 *citra test*. Walaupun hasil test independen kini sudah tersedia, ukuran sampel per kelas masih cukup kecil untuk menarik kesimpulan final mengenai robustitas model di berbagai kondisi lapangan. Validasi silang, perluasan dataset, dan pembandingan dengan arsitektur lain tetap diperlukan agar klaim performa menjadi lebih kuat. Fine-tuning dari layer ke-80 juga masih perlu diuji melalui studi ablasi, misalnya dengan membandingkan titik unfreeze ke-60, ke-100, dan ke-120.

Perbedaan antara *macro precision* 0,8170 dan *macro recall* 0,7875 pada *test* independen memberi petunjuk penting. *Precision* yang lebih tinggi menunjukkan bahwa saat model memutuskan sebuah kelas, prediksi tersebut relatif tepat. *Recall* yang lebih rendah menunjukkan bahwa sebagian sampel dari beberapa kelas masih terlewat. Masalah utama bukan hanya kesalahan acak, melainkan sensitivitas kelas yang belum merata. Kelas dengan *F1-score* rendah seperti Papua_Tifa dan Aceh_Pintu_Aceh perlu menjadi prioritas untuk penambahan data dan inspeksi visual lanjutan.

Implementasi Flask memperlihatkan potensi penerapan model dalam sistem interaktif. Namun, pembahasan aplikasi harus ditempatkan secara proporsional. Sebelum dipakai di lingkungan nyata, sistem perlu diuji pada citra batik dari kondisi lapangan, diberi validasi keamanan unggahan, dilengkapi pencatatan prediksi, dan diukur waktu inferensinya. Dengan demikian, aplikasi web merupakan keluaran tambahan yang mendukung demonstrasi, bukan pengganti validasi model.

4. PENUTUP

Simpulan dan Saran

Penelitian ini menjawab rumusan masalah sebagai berikut. Pertama, pipeline klasifikasi 20 motif batik Nusantara berhasil dirancang dengan EfficientNetB0, *preprocessing* 224 x 224 piksel, augmentasi numerik, pelatihan dua tahap, dan integrasi Flask. Kedua, model memperoleh *validation accuracy* 81,25% pada 80 citra validasi, lalu menghasilkan akurasi 78,75%, *macro precision* 0,8170, *macro recall* 0,7875, *macro F1-score* 0,7843, dan *test loss* 1,1609 pada 160 citra test independen. Ketiga, kelas Jawa_Barat_Megamendung, Maluku_Pala, dan Yogyakarta_Kawung menunjukkan performa tertinggi, sedangkan Papua_Tifa, Aceh_Pintu_Aceh, Jawa_Timur_Pring, NTB_Lumbung, dan Yogyakarta_Parang memerlukan perhatian lebih karena performanya lebih rendah pada test independen. Hasil ini mendukung penggunaan EfficientNetB0 sebagai baseline awal yang cukup stabil, tetapi belum membuktikan keunggulan arsitektur secara umum. Penelitian berikutnya perlu memperluas dataset, melakukan validasi silang, membandingkan EfficientNetB0 dengan arsitektur lain, menguji aplikasi pada citra batik dari kondisi nyata, dan melakukan studi ablasi pada strategi *fine-tuning*.

DAFTAR PUSTAKA

Quan, H., Li, Y., Liu, D., & Zhou, Y. (2024). Protection of Guizhou Miao Batik Culture Based on Knowledge Graph and Deep Learning. arXiv. https://doi.org/10.48550/arXiv.2404.06168

Shaharuddin, S. I. S., Shamsuddin, M. S., Drahman, M. H., Hasan, Z., & Mohd Asri, N. A. (2021). A Review on the Malaysian and Indonesian Batik Production, Challenges, and Innovations in the 21st Century. SAGE Open.

Hajura, S. (2022). Shared Heritage Diplomacy of Indonesia and Malaysia as Soft Power in the Southeast Asia Region. Jurnal Hubungan Internasional.

Handayani, W., Kristijanto, A. I., & Hunga, A. I. R. (2018). Behind the eco-friendliness of "batik warna alam": Discovering the motives behind the production of batik in Jarum village, Klaten. Wacana, Journal of the Humanities of Indonesia.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11976-11986.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, 6105-6114.

Bruno, A., Moroni, D., & Martinelli, M. (2025). Efficient Adaptive Ensembling for Image Classification. *Expert Systems*, *42*(1), e13424. https://doi.org/10.1111/exsy.13424

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2023). A Survey on Visual Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 87–110. https://doi.org/10.1109/TPAMI.2022.3152247

Hasana, M. M., Ibrahim, M., & Ali, M. S. (2023). *Speeding Up EfficientNet: Selecting Update Blocks of Convolutional Neural Networks using Genetic Algorithm in Transfer Learning* (arXiv:2303.00261). arXiv. https://doi.org/10.48550/arXiv.2303.00261

Jiang, J., Shu, Y., Wang, J., & Long, M. (2022). *Transferability in Deep Learning: A Survey* (arXiv:2201.05867). arXiv. https://doi.org/10.48550/arXiv.2201.05867

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in Vision: A Survey. *ACM Computing Surveys*, *54*(10s), 1–41. https://doi.org/10.1145/3505244

Kumar, T., Mileo, A., Brennan, R., & Bendechache, M. (2023). *Image Data Augmentation Approaches: A Comprehensive Survey and Future directions* (arXiv:2301.02830). arXiv. https://doi.org/10.48550/arXiv.2301.02830

Octadion, O., Yudistira, N., & Kurnianingtyas, D. (2023). *Synthesis of Batik Motifs using a Diffusion—Generative Adversarial Network* (arXiv:2307.12122). arXiv. https://doi.org/10.48550/arXiv.2307.12122

Plested, J., Phiri, M., & Gedeon, T. (2025). *Deep transfer learning for image classification: A survey* (arXiv:2205.09904). arXiv. https://doi.org/10.48550/arXiv.2205.09904

Sapkota, R., Raza, S., Shoman, M., Paudel, A., & Karkee, M. (2025). *Multimodal Large Language Models for Image, Text, and Speech Data Augmentation: A Survey* (arXiv:2501.18648). arXiv. https://doi.org/10.48550/arXiv.2501.18648

Wang, Y., Zhang, Z.-H., Xu, S. X., & Guo, W. (2024). *Generalizing Few Data to Unseen Domains Flexibly Based on Label Smoothing Integrated with Distributionally Robust Optimization* (arXiv:2408.05082). arXiv. https://doi.org/10.48550/arXiv.2408.05082

Wei, J., Torresani, L., Wei, J., & Hassanpour, S. (2022). *Calibrating Histopathology Image Classifiers using Label Smoothing* (arXiv:2201.11866). arXiv. https://doi.org/10.48550/arXiv.2201.11866

Yang, Z., Sinnott, R. O., Bailey, J., & Ke, Q. (2022). *A Survey of Automated Data Augmentation Algorithms for Deep Learning-based Image Classification Tasks* (arXiv:2206.06544). arXiv. https://doi.org/10.48550/arXiv.2206.06544