

Klasifikasi Motif Batik Nusantara Menggunakan Vision Transformer (ViT) Berbasis Deep Learning

Imam Fathurrahman^{1*}, Muhammad Djamaluddin², Zaenul Amri³, M. Nurul Wathani⁴

^{1,2,3,4} Program Studi Informatika, Universitas Hamzanwadi

*har.poenya@gmail.com

Abstrak

Batik merupakan salah satu warisan budaya Indonesia yang mengandung nilai filosofi dan identitas lokal dari berbagai daerah. Di era digital, klasifikasi motif batik secara otomatis menjadi penting untuk mendukung pelestarian, edukasi, dan komersialisasi budaya. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi motif batik nusantara menggunakan arsitektur Vision Transformer (ViT), sebuah pendekatan deep learning berbasis self-attention yang mampu memahami hubungan spasial global dalam citra. Dataset yang digunakan terdiri dari 800 citra dengan 20 kelas motif batik dari berbagai daerah, yang dibagi ke dalam data latih dan uji. Model ViT dilatih menggunakan pretrained weights dari ImageNet-21k, dengan preprocessing standar dan augmentasi pada data latih. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, F1-score, confusion matrix, dan visualisasi prediksi. Hasil menunjukkan bahwa model ViT mampu mencapai akurasi hingga 96%, dengan sebagian besar kelas memperoleh nilai F1-score di atas 0.90. Pengujian pada citra baru yang tidak ada dalam dataset pelatihan menunjukkan kemampuan generalisasi yang baik dengan confidence 99.94%. Temuan ini membuktikan bahwa ViT merupakan arsitektur yang efektif dan efisien untuk tugas klasifikasi motif batik, sekaligus berkontribusi pada pelestarian budaya melalui teknologi kecerdasan buatan.

Kata kunci : Batik, Vision Transformer, Deep Learning, Klasifikasi Citra, Pelestarian Budaya

Abstract

Batik is a cultural heritage of Indonesia that reflects local philosophies and identities through its diverse motifs. In the digital era, automatic classification of batik patterns plays a crucial role in cultural preservation, education, and commercialization. This study aims to develop a batik motif classification system using Vision Transformer (ViT), a deep learning architecture based on self-attention capable of capturing global spatial relationships in images. The dataset comprises 800 images spanning 20 batik motif classes from various regions, divided into training and testing subsets. The ViT model was fine-tuned using pretrained weights from ImageNet-21k, with standard preprocessing and data augmentation applied to the training set. Model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix, and prediction visualization. Results indicate that ViT achieved an overall accuracy of 96%, with most classes recording F1-scores above 0.90. Evaluation on unseen batik images demonstrated excellent generalization capability, achieving 99.94% confidence in prediction. These findings suggest that ViT is an effective and efficient architecture for batik motif classification and offers valuable contributions to cultural preservation through artificial intelligence.

Keywords : Batik, Vision Transformer, Deep Learning, Image Classification, Cultural Preservation.

1. Pendahuluan

Batik adalah warisan budaya takbenda Indonesia yang diakui UNESCO sejak 2009 sebagai Masterpiece of the Oral and Intangible Heritage of Humanity. Setiap motif batik mengandung nilai

filosofis, simbolis, dan identitas lokal. Sebagai contoh motif *Parang* dari Yogyakarta melambangkan keteguhan, sedangkan *Mega Mendung* dari Cirebon merepresentasikan kesejukan. Di era digital, pelestarian batik

menghadapi tantangan akibat perubahan generasi dan gaya hidup. Oleh sebab itu, pemanfaatan teknologi digital, khususnya kecerdasan buatan, menjadi strategi penting untuk mendukung pelestarian dan klasifikasi motif batik berbasis teknologi.

Klasifikasi motif batik secara otomatis menggunakan metode *computer vision* dan *deep learning* telah menjadi topik yang semakin berkembang dalam dekade terakhir. Salah satu pendekatan paling populer adalah *Convolutional Neural Network* (CNN), yang dikenal efektif dalam mengekstraksi fitur lokal dari citra visual. CNN telah digunakan dalam berbagai studi klasifikasi batik dan memberikan hasil yang signifikan. Misalnya, penelitian oleh Wona *et al.* menunjukkan bahwa model CNN Hasil pengujian menunjukkan akurasi sebesar 91,24% pada data testing, membuktikan bahwa model CNN efektif untuk klasifikasi citra batik [1]. Penelitian lain oleh Wibowo *et al.* dan Negara *et al.* juga menunjukkan bahwa teknik *transfer learning* berbasis arsitektur CNN seperti *VGG-16* dan *ResNet-50* secara signifikan meningkatkan akurasi pada dataset motif batik yang kompleks [2],[3].

Meskipun CNN unggul dalam ekstraksi fitur lokal, keterbatasannya dalam memahami hubungan spasial global mendorong munculnya *Vision Transformer* (ViT) oleh Dosovitskiy *et al.*, yang menggunakan mekanisme *self-attention* untuk memahami pola visual secara menyeluruh [4]. ViT

terbukti unggul dalam berbagai tugas klasifikasi citra seperti *ImageNet*. Dalam konteks batik, Achmadiyah *et al.* menunjukkan bahwa model *Multi-Stage ViT* (MSViT) lebih akurat dan efisien dibanding CNN [5]. Di sisi lain, penelitian oleh Abd Manap *et al.* menunjukkan bahwa arsitektur CNN modern seperti YOLOv8 masih sangat kompetitif, dengan akurasi hingga 98,8% dalam klasifikasi motif batik Malaysia dan Indonesia, menunjukkan bahwa pilihan arsitektur tetap krusial, meski tidak berbasis ViT [6].

Salah satu keunggulan ViT adalah kemampuannya untuk ditingkatkan melalui pendekatan *transfer learning*, yaitu dengan memanfaatkan model ViT yang telah dilatih sebelumnya (*pretrained*) pada dataset besar seperti *ImageNet-21k*. Pendekatan ini memungkinkan peningkatan akurasi walaupun digunakan pada dataset batik yang terbatas, serta mempercepat proses pelatihan tanpa memerlukan infrastruktur komputasi yang berlebihan.

Dengan mempertimbangkan kompleksitas visual motif batik serta keterbatasan jumlah data yang tersedia secara publik, ViT memberikan solusi yang lebih kontekstual dan adaptif dibandingkan pendekatan konvensional. Oleh karena itu, penelitian ini bertujuan untuk menerapkan dan mengevaluasi performa *Vision Transformer* dalam klasifikasi motif batik nusantara. Diharapkan, pendekatan ini dapat menghasilkan

sistem klasifikasi citra yang akurat, efisien, dan berkontribusi pada pelestarian budaya Indonesia di era digital

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Beberapa penelitian terdahulu telah memberikan kontribusi penting dalam pengembangan sistem klasifikasi motif batik berbasis *deep learning*.

Salah satunya adalah studi yang dilakukan oleh :

- Elvitaria *et al.* yang menunjukkan bahwa pendekatan *transfer learning* dengan memanfaatkan arsitektur CNN *pretrained* seperti *MobileNetV3* mampu meningkatkan akurasi klasifikasi batik, khususnya pada dataset berukuran kecil. Selain itu, pendekatan ini juga mempercepat proses pelatihan dan mendukung inisiatif *digitalisasi* budaya berbasis kecerdasan buatan [7].
- Pengembangan arsitektur model yang lebih kompleks ditunjukkan oleh Achmadiyah *et al.*, yang merancang *Multi-Stage Vision Transformer* (MSViT) untuk tugas klasifikasi motif batik. Model ini mengintegrasikan teknik konvolusi sebagai *downsampling* dan *Vanilla Attention* sebagai *token mixer*. Pengujian pada dua dataset berbeda, yaitu Batik 300 dan Batik Nitik 960, menunjukkan bahwa MSViT mengungguli model CNN baik dari segi akurasi, efisiensi komputasi, maupun jumlah parameter yang digunakan [5].

- Penelitian oleh Wona *et al* membuktikan bahwa sistem klasifikasi batik Indonesia berbasis web dengan CNN mampu mengenali 14 motif dari 3.880 citra dengan akurasi 91,24%, sehingga efektif mendukung pelestarian dan promosi batik secara digital berbasis teknologi [1].
- Selain itu, Wibowo *et al.* membuktikan keunggulan *transfer learning* dalam klasifikasi motif batik. Mereka membandingkan beberapa arsitektur seperti CNN biasa, VGG-16, dan *Xception* untuk lima jenis motif batik, dan model terbaik (CNN dan VGG16) berhasil mencapai akurasi pelatihan 91,23% serta akurasi pengujian 89%, menandakan peningkatan performa yang signifikan melalui pemanfaatan model *pra-latih* [2].
- Dalam penelitian lain, Negara *et al.* menggunakan *ResNet-50* yang dipadukan dengan teknik *augmentasi* data. Hasil klasifikasi menunjukkan akurasi hingga 96%, mengonfirmasi bahwa *ResNet* mampu mengenali pola visual yang kompleks dalam motif batik dengan sangat baik [3].
- Terakhir, Sugiarto *et al.* dan Wicaksono *et al.* menekankan bahwa meskipun CNN dapat mencapai akurasi tinggi dalam pengenalan motif batik, keterbatasannya terletak pada kemampuan memahami konteks global dalam citra, terutama pada motif batik yang

memiliki pola menyebar dan tidak beraturan [8], [9]

2.2. Landasan Teori

1. Batik Nusantara

Indonesia dikenal sebagai negara kepulauan dengan keanekaragaman etnis, budaya, dan bahasa yang sangat kaya. Salah satu unsur budaya yang menjadi simbol kuat identitas nasional adalah kain batik, yang tersebar luas di berbagai daerah di tanah air. Nilai budaya batik telah mendapat apresiasi, tidak hanya dari masyarakat Indonesia sendiri, tetapi juga dari komunitas internasional. Pada 2 Oktober 2009, UNESCO menetapkan batik Indonesia sebagai *Intangible Cultural Heritage of Humanity* dengan memasukkannya ke dalam daftar representatif warisan budaya takbenda dunia. Sejak saat itu, tanggal tersebut diperingati secara nasional sebagai Hari Batik [10], [11].

2. Klasifikasi Citra dan Deep learning

Klasifikasi citra merupakan proses penentuan label atau kategori pada sebuah citra berdasarkan ciri visual yang terkandung di dalamnya, seperti warna, bentuk, tekstur, dan struktur spasial. Tujuan utama dari proses ini adalah untuk mengelompokkan citra ke dalam kelas-kelas yang telah ditentukan sebelumnya, sehingga sistem komputer mampu memahami dan menganalisis isi citra secara otomatis [12], [13].

Dalam beberapa tahun terakhir, pendekatan *deep learning* telah merevolusi cara sistem memproses citra digital. Tidak seperti pendekatan konvensional yang membutuhkan ekstraksi fitur secara manual, *deep learning* menggunakan jaringan saraf berlapis (*deep neural networks*) untuk mempelajari representasi fitur secara otomatis langsung dari data mentah. Jaringan ini mempelajari fitur-fitur secara bertingkat dari representasi visual tingkat rendah hingga tingkat tinggi melalui proses pelatihan *end-to-end* yang efisien [14].

Efektivitas pendekatan ini telah dibuktikan dalam berbagai studi dan kompetisi global, di mana model berbasis *deep learning* menunjukkan performa yang unggul dalam mengenali pola visual kompleks, khususnya pada dataset berskala besar [15].

3. Vision Transformer (ViT)

Arsitektur ViT pertama kali diperkenalkan oleh Dosovitskiy *et al.* sebagai pendekatan alternatif terhadap model berbasis konvolusi dalam tugas pengolahan citra. ViT mengadaptasi arsitektur *Transformer*, yang sebelumnya telah berhasil merevolusi bidang *Natural Language Processing* (NLP), untuk digunakan dalam domain visual. Pendekatan ini memproses citra sebagai urutan tambalan (*image patches*) yang dikodekan dan diperlakukan serupa dengan token dalam teks, sehingga memungkinkan penerapan mekanisme

self-attention untuk menangkap hubungan spasial secara menyeluruh dalam citra [4].

Berbeda dengan *Convolutional Neural Networks* (CNN) yang menggunakan kernel *konvolusi lokal* untuk mengekstraksi fitur spasial, ViT secara langsung memodelkan *dependensi global* melalui *attention*, tanpa asumsi lokalitas atau *translasi invarian*. Hal ini membuat ViT lebih fleksibel dalam mengenali pola-pola kompleks yang tidak selalu memiliki struktur spasial yang teratur.

Beberapa keunggulan utama dari *Vision Transformer* adalah (1) Kemampuan menangkap *dependensi global*, yang memungkinkan pemahaman konteks visual secara menyeluruh. (2) Arsitektur *non-konvolusional*, yang memungkinkan fleksibilitas lebih tinggi terhadap variasi pola dan struktur data visual. (3) Performa kompetitif pada benchmark skala besar, seperti *ImageNet-1K* dan *ImageNet-21K*, terutama ketika dilatih dengan data berukuran besar.

Meskipun demikian, performa ViT sangat bergantung pada ketersediaan data pelatihan dalam jumlah besar dan teknik regularisasi yang tepat. Studi lanjutan menunjukkan bahwa dengan pelatihan dan *augmentasi* yang memadai, ViT dapat menyamai bahkan melampaui CNN pada banyak klasifikasi citra [16].

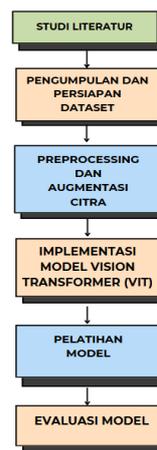
4. Citra Digital

Citra (*image*) merupakan representasi dua dimensi dari suatu objek, yang tersusun atas elemen-elemen citra yang disebut piksel (*picture*

elements) [18], [19]. Citra digital dapat dinyatakan sebagai matriks dua dimensi yang terdiri dari baris dan kolom, di mana setiap posisi indeks menyatakan koordinat spasial (x, y), dan setiap nilai dalam matriks merepresentasikan tingkat intensitas atau kecerahan pada titik tersebut [13], [17]. Menurut Gonzalez dan Woods, citra digital didefinisikan sebagai fungsi dua variabel $f(x,y)$, di mana x dan y menunjukkan koordinat spasial, dan nilai f pada titik tersebut menyatakan intensitas atau tingkat keabuan (*gray level*) dari citra. Titik-titik ini secara keseluruhan membentuk representasi visual dari objek yang diamati [20].

2.3. Tahapan Penelitian Menggunakan ViT

Tahapan-tahapan dalam penelitian ini dapat digambarkan secara visual pada Gambar 1.



Gambar 1. Tahapan Penelitian menggunakan ViT

1. Studi Literatur

Penelitian diawali dengan studi literatur yang bertujuan memahami teori dasar batik sebagai warisan budaya Indonesia, konsep pengolahan

citra digital, serta perkembangan arsitektur *deep learning*, khususnya ViT. Kajian ini juga mencakup telaah terhadap penelitian-penelitian terdahulu untuk mengidentifikasi kesenjangan (*gap*) dan justifikasi pemilihan metode yang digunakan.

2. Pengumpulan dan Persiapan Dataset

Langkah berikutnya adalah mengumpulkan dan menyiapkan data citra batik dari berbagai sumber yang relevan. Data tersebut kemudian diklasifikasikan dan dipersiapkan agar siap digunakan dalam proses analisis lebih lanjut.

3. Preprocessing dan Augmentasi Citra

Data citra yang telah dikumpulkan selanjutnya diproses dan disesuaikan agar memenuhi kebutuhan penelitian, termasuk melalui tahapan normalisasi dan *augmentasi* untuk meningkatkan kualitas dan variasi data.

4. Implementasi Model ViT

Setelah data siap, dilakukan implementasi model berbasis *deep learning* yang sesuai dengan kebutuhan klasifikasi motif batik. Pada tahap ini, model ViT dipilih dikembangkan atau dimodifikasi agar dapat mengenali pola pada data citra secara optimal.

5. Pelatihan Model

Model yang telah diimplementasikan kemudian dilatih menggunakan data yang sudah diproses, dengan memperhatikan parameter dan strategi pelatihan yang telah ditentukan sebelumnya.

6. Evaluasi Model

Tahap terakhir adalah mengevaluasi performa model yang telah dilatih. Evaluasi dilakukan menggunakan berbagai metrik yang umum digunakan dalam klasifikasi citra, serta dapat disertai analisis visual dan pengukuran tingkat keyakinan model terhadap hasil prediksi.

3. Metode Penelitian

3.1. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan kumpulan citra digital motif batik dari *Kaggle*, yang diterbitkan oleh HendryHB dengan *lisensi CC0: Public Domain* [21]. Dataset terdiri atas 800 citra motif batik, serta terbagi menjadi dua subset: 640 citra (80%) untuk data latih (*train*) dan 160 citra (20%) untuk data uji (*test*) sesuai struktur aslinya. Seluruh data telah diverifikasi dari segi label dan kualitas citra sebelum digunakan dalam proses pelatihan dan evaluasi model.

3.2. Instrumen Penelitian

Instrumen penelitian ini terdiri dari perangkat lunak berupa Python beserta beberapa library pendukung, seperti *timm*, *torch (PyTorch)*, *scikit-learn*, *matplotlib*, dan *seaborn*. Seluruh perangkat lunak ini digunakan untuk proses implementasi, pelatihan, evaluasi, serta visualisasi model *deep learning* yang dikembangkan dalam penelitian ini.

3.3. Teknik Analisis Data

Analisis data dilakukan secara kuantitatif dan kualitatif.

1. *Preprocessing* dan *augmentasi* citra

Seluruh citra pada data latih diproses melalui tahap *resize*, *normalisasi* (menggunakan nilai *mean* dan *standar deviasi* dari *ImageNet*), serta *augmentasi* seperti *random rotation*, *horizontal flip*, dan *color jitter* untuk meningkatkan variasi dan generalisasi model. Data uji hanya melalui tahap *resize* dan *normalisasi*, tanpa *augmentasi*, agar evaluasi objektif.

2. Implementasi model

- Penelitian ini menggunakan arsitektur ViT, tepatnya model *vit-base-patch16-224*, yang dimanfaatkan melalui *library timm* dan *pretrained* pada *ImageNet-21k*.
- Model dimodifikasi pada output layer agar sesuai dengan jumlah kelas motif batik.

3. Pelatihan model

Model dilatih menggunakan fungsi *loss CrossEntropyLoss* dan *optimizer AdamW* dengan *learning rate* 0.0001, *batch size* 16, dan jumlah *epoch* 50. Selama pelatihan, performa model divalidasi secara periodik untuk memilih hasil terbaik.

4. Evaluasi model

Performa model diuji menggunakan data test dan dievaluasi menggunakan beberapa metrik:

- *Classification Report*, yang memuat detail evaluasi per-kelas motif, terdiri dari :
 - Akurasi, yang mengukur proporsi total prediksi yang benar terhadap seluruh data uji.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

- *Precision*, yang menilai ketepatan prediksi positif per kelas

$$Precision = \frac{TP}{(TP) + (FP)} \times 100\% \quad (2)$$

- *Recall*, yang menilai sejauh mana model berhasil mengidentifikasi seluruh sampel aktual dari setiap kelas.

$$Recall = \frac{TP}{(TP)+(FN)} \times 100\% \quad (3)$$

- *F1-Score*, yang merupakan rata-rata harmonis dari *precision* dan *recall*.

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} \quad (4)$$

- *Confusion Matrix*, digunakan untuk memetakan jumlah prediksi benar dan salah antar kelas. Matriks ini berguna untuk mengidentifikasi kelas motif batik yang paling sering tertukar, khususnya kelas-kelas dengan kemiripan pola atau warna.
- Visualisasi Prediksi Sampel dimana beberapa citra dari data uji ditampilkan bersama hasil prediksinya untuk melihat secara langsung ketepatan model dalam mengenali motif. Visualisasi ini penting sebagai evaluasi kualitatif, terutama pada motif-motif yang kompleks.
- *Confidence Score* yaitu tingkat keyakinan model terhadap prediksi, dengan memberikan gambaran seberapa yakin model dalam menetapkan sebuah citra ke salah satu kelas tertentu. Nilai ini berguna untuk mengevaluasi kepercayaan model dalam kondisi ambigu.

3.4. Lokasi Penelitian

Penelitian dilaksanakan di Laboratorium Informatika, Fakultas Teknik, Universitas Hamzanwadi selama Maret – Mei 2025.

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

1. Persiapan Dataset

Dataset yang digunakan dalam format terkompresi (*.zip). Setelah proses ekstraksi, dataset tersebut terbagi ke dalam dua direktori utama, yaitu folder *train* untuk pelatihan model dan folder *test* untuk evaluasi performa.

Folder *train* berisi total 640 citra yang mewakili 20 kategori motif batik nusantara, dengan masing-masing kategori terdiri dari 32 citra. Adapun kategori tersebut terlihat pada tabel 1.

Tabel 1. Daftar Kategori Motif Batik

| Aceh_Pintu_Aceh | Bali_Barong | Bali_Merak | DKI_Ondel_Ondel |
|----------------------------|-------------------------|-------------------|-------------------------|
| JawaBarat_Megamendung | JawaTimur_Pring | Kalimantan_Dayak | Lampung_Gajah |
| Madura_Mataketeran | Maluku_Pala | NTB_Lumpang | Papua_Asmat |
| Papua_Cendrawasih | Papua_Tifa | Solo_Parang | SulawesiSelatan_Lontara |
| SumateraBarat_Rumah_Minang | SumateraUtara_Boraspati | Yogyakarta_Kawung | Yogyakarta_Parang |

Sementara itu, folder *test* berisi 160 citra yang terdiri dari 8 citra per-kategori, dengan distribusi label yang sama dengan folder *train*. Dengan demikian, total citra dalam dataset berjumlah 800 citra. Setiap citra telah dikategorikan berdasarkan nama daerah dan jenis motif batik yang diwakilinya. Kategori-kategori ini kemudian dijadikan sebagai label kelas, yang berfungsi

sebagai target utama dalam proses pelatihan model klasifikasi. Struktur label yang tersusun secara sistematis mempermudah proses pengelompokan data serta evaluasi performa model terhadap masing-masing kelas. Representasi visual dari distribusi kategori dan label kelas dalam dataset disajikan antara lain.

```
Jumlah gambar di training: 640
Jumlah gambar di testing: 160
Jumlah kelas: 20
Label: ['Aceh_Pintu_Aceh', 'Bali_Barong', 'Bali_Merak',
```

Gambar 2. Pembagian Dataset

2. Preprocessing dan Augmentasi Citra

Di awal proses perapihan dataset, seluruh citra dalam folder *train* dan *test* terlebih dahulu diperiksa konsistensi format dan dimensi citranya. Citra yang memiliki resolusi tidak seragam atau format ekstensi yang tidak didukung (misalnya *.png di antara *.jpg) diubah agar sesuai standar input model, yaitu berukuran 224 × 224 piksel dan berekstensi *.jpg Ilustrasi proses ini ditampilkan pada gambar 3. Kemudian dilakukan *augmentasi* pada dataset *training*.



Gambar 3. Citra Batik Nusantara

3. Implementasi Model Vision Transformer

Penelitian ini menggunakan arsitektur *Vision Transformer* (ViT), tepatnya varian *vit_base_patch16_224*, sebagai model utama

dalam klasifikasi motif batik. Model ini dipilih karena kemampuannya dalam memahami struktur visual secara *global* melalui mekanisme *self-attention*. Implementasi dilakukan dengan memanfaatkan model *pretrained* dari *library timm*, yang telah dilatih pada dataset *ImageNet-21k*, sehingga memungkinkan pendekatan *transfer learning* untuk meningkatkan efisiensi pelatihan dan akurasi pada dataset batik yang relatif terbatas. Proses ini digambarkan secara visual pada Gambar 4.

```
import timm
model = timm.create_model('vit_base_patch16_224', pretrained=True)
model.head = nn.Linear(model.head.in_features, num_classes)
model.to(device)
```

Gambar 4. Penggunaan *Library Timm*

Proses pelatihan dilakukan selama 50 *epoch* dengan *batch size* 16, menggunakan *optimizer Adam*. Selama pelatihan, performa model dimonitor menggunakan akurasi validasi untuk memastikan proses konvergen.

4. Evaluasi Model

Untuk mengevaluasi performa model secara menyeluruh, dilakukan kombinasi evaluasi kuantitatif melalui metrik statistik dan evaluasi kualitatif melalui visualisasi prediksi citra. Berikut pembahasannya berdasarkan empat visual utama.

5. *Classification Report*

Hasil evaluasi performa model terlihat pada Gambar 5, yang menampilkan *classification report* dari hasil pengujian terhadap data uji. Secara keseluruhan, model menunjukkan bahwa model

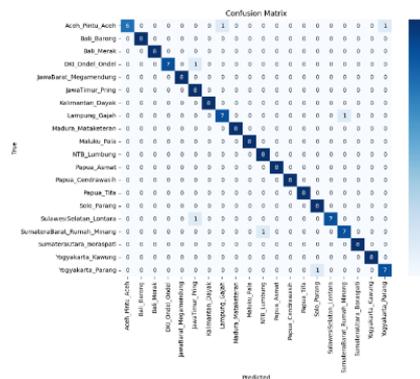
ViT mencapai akurasi keseluruhan sebesar 96%. Nilai *precision*, *recall*, dan *F1-score* untuk sebagian besar kelas berada di atas 0.90, bahkan banyak di antaranya mencapai nilai sempurna (1.00). Ini mengindikasikan model mampu mengenali motif-motif batik dari berbagai daerah secara akurat dan konsisten. Namun, beberapa kelas seperti *Aceh_Pintu_Aceh* memiliki nilai *recall* lebih rendah (0.75), yang menunjukkan bahwa sebagian prediksi dari kelas ini salah diklasifikasikan ke kelas lain.

Classification Report:

| | precision | recall | f1-score | support |
|----------------------------|-----------|--------|----------|---------|
| Aceh_Pintu_Aceh | 1.00 | 0.75 | 0.85 | 8 |
| Bali_Barong | 1.00 | 1.00 | 1.00 | 8 |
| Bali_Merak | 1.00 | 1.00 | 1.00 | 8 |
| DKI_Ondel_Ondel | 1.00 | 0.88 | 0.93 | 8 |
| JawaBarat_Megamendung | 1.00 | 1.00 | 1.00 | 8 |
| JawaTimur_Pring | 0.80 | 1.00 | 0.89 | 8 |
| Kalimantan_Dayak | 1.00 | 1.00 | 1.00 | 8 |
| Lampung_Gajah | 0.88 | 0.88 | 0.88 | 8 |
| Madura_Rataketeran | 1.00 | 1.00 | 1.00 | 8 |
| Maluku_Pala | 1.00 | 1.00 | 1.00 | 8 |
| NTB_Lumbang | 0.89 | 1.00 | 0.94 | 8 |
| Papua_Asmat | 1.00 | 1.00 | 1.00 | 8 |
| Papua_Cendrawasih | 1.00 | 1.00 | 1.00 | 8 |
| Papua_Tifa | 1.00 | 1.00 | 1.00 | 8 |
| Solo_Parang | 0.89 | 1.00 | 0.94 | 8 |
| SulawesiSelatan_Lontara | 1.00 | 0.88 | 0.93 | 8 |
| SumateraBarat_Rumah_Minang | 0.88 | 0.88 | 0.88 | 8 |
| SumateraUtara_Boraspati | 1.00 | 1.00 | 1.00 | 8 |
| Yogyakarta_Kawang | 1.00 | 1.00 | 1.00 | 8 |
| Yogyakarta_Parang | 0.88 | 0.88 | 0.88 | 8 |
| accuracy | | | 0.96 | 160 |
| macro avg | 0.96 | 0.96 | 0.96 | 160 |
| weighted avg | 0.96 | 0.96 | 0.96 | 160 |

Gambar 5. *Classification Report*

6. Confusion Matrix



Gambar 6. *Confusion matrix*

Confusion matrix yang ditampilkan pada Gambar 6 memperlihatkan performa prediksi per kelas

secara visual. Terlihat bahwa mayoritas prediksi berada pada diagonal utama, menandakan klasifikasi yang tepat. Namun, terjadi kesalahan klasifikasi kecil, seperti:

- *Aceh_Pintu_Aceh* salah diklasifikasikan sebagai *Yogyakarta_Parang* dan *Lampung_Gajah*.
- *DKI_Ondel_Ondel* keliru satu kali sebagai *JawaBarat_Megamendung*.

Kesalahan-kesalahan ini biasanya disebabkan oleh kemiripan elemen visual, seperti bentuk pola atau palet warna.

7. Visualisasi Prediksi Sampel



Gambar 7. Hasil prediksi sampel

Gambar 7 menampilkan hasil visualisasi prediksi terhadap enam citra dari kelas *Aceh_Pintu_Aceh* menunjukkan bahwa model mengenali 4 citra dengan benar, sementara 2 citra batik salah diklasifikasikan ke kelas lain. Misalnya, satu citra salah diprediksi sebagai *Yogyakarta_Parang*, dan satu lagi sebagai *Lampung_Gajah*. Hal ini menguatkan temuan pada *confusion matrix*, bahwa motif dengan pola simetris dan warna dominan merah/coklat dapat membingungkan model dalam membedakan antar kelas.

8. Confidence Score



Gambar 8. Confidence Score

Gambar 8 menampilkan hasil pengujian model Untuk menguji kemampuan generalisasi model, sebuah citra batik motif *Ondel- ondel* dari DKI jakarta yang tidak termasuk dalam dataset dilibatkan dalam prediksi. Hasilnya, model mengklasifikasikan citra tersebut sebagai *DKI_Ondel_Ondel* dengan tingkat kepercayaan (*confidence*) 99.94%. Ini membuktikan bahwa ViT mampu mengenali motif baru dengan akurasi tinggi, selama struktur visualnya masih merepresentasikan pola kelas yang telah dipelajari.

4.2. Pembahasan

Penelitian ini membuktikan bahwa *Vision Transformer (ViT)* sangat baik untuk mengenali dan membedakan motif batik dari berbagai daerah di Indonesia. Model ini bisa memprediksi motif dengan akurasi sangat tinggi, bahkan untuk motif yang rumit. Beberapa motif seperti *Bali_Barong* dan *Papua_Tifa* dikenali dengan benar.

Beberapa motif batik memang memiliki kemiripan visual, sehingga model terkadang salah mengklasifikasikan, misalnya motif

Aceh_Pintu_Aceh yang kadang dikenali sebagai *Yogyakarta_Parang* atau *Lampung_Gajah* karena kesamaan bentuk atau warna. Namun, pengujian lebih lanjut menunjukkan bahwa ViT mampu mengenali motif baru seperti *DKI_Ondel_Ondel* yang belum pernah dilatih sebelumnya, dengan tingkat kepercayaan hingga 99,94%.

Berdasarkan keseluruhan temuan, ViT terbukti menjadi arsitektur yang sangat kompetitif untuk klasifikasi batik, karena kemampuannya memahami konteks visual secara menyeluruh.

5. Kesimpulan

Penelitian ini menunjukkan bahwa *Vision Transformer* (ViT) mampu mengklasifikasikan motif batik nusantara dengan akurasi tinggi, mencapai 96%. Keunggulan ViT terletak pada kemampuannya memahami pola visual global. Model juga berhasil menggeneralisasi terhadap citra baru yang tidak terdapat dalam dataset pelatihan, menandakan potensinya dalam aplikasi nyata seperti digitalisasi budaya dan edukasi visual. Kesalahan klasifikasi yang terjadi disebabkan oleh kemiripan antar motif, namun secara keseluruhan, ViT terbukti efektif dan efisien dalam mengenali kompleksitas motif batik. Penggunaan *transfer learning* juga terbukti mampu mengoptimalkan kinerja model meskipun dengan jumlah data terbatas. ViT menjadi pendekatan yang menjanjikan dalam pelestarian budaya melalui teknologi kecerdasan buatan.

6. Daftar Pustaka

- [1] M. M. A. Wona *et al.*, "Klasifikasi Batik Indonesia Menggunakan Convolutional Neural Network (CNN)," *Jurnal Rekayasa Teknologi Informasi (JURTI)*, vol. 7, no. 2, p. 172, Dec. 2023, doi: 10.30872/jurti.v7i2.13694.
- [2] M. Wibowo, P. Paradise, and R. F. Alya, "Classification Of Batik Motif Using Transfer Learning On Convolutional Neural Network (CNN)," *Jurnal Teknik Informatika (Jutif)*, 2023, doi: 10.52436/1.jutif.2023.4.1.564.
- [3] B. S. Negara, E. Satria, S. Sanjaya, and D. R. D. Santoso, "ResNet-50 for Classifying Indonesian Batik with Data Augmentation," *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, pp. 1–4, 2021, doi: 10.1109/ICOTEN52080.2021.9493488.
- [4] A. Dosovitskiy *et al.*, "an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.
- [5] M. N. Achmadiyah, C.-C. Sun, W. Kuo, and N. Setyawan, "Multi-Stage Vision Transformer for Batik Classification," *2024 International Electronics Symposium (IES)*, pp. 449–453, 2024, doi: 10.1109/ies63037.2024.10665807.
- [6] N. Abd Manap, L. Xiao Xuan, K. Kumar Singh, A. Sheikh Akbari, and A. Putra, "Classification of Malaysian and Indonesian Batik Designs Using Deep Learning Models," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 16, no. 4, pp. 23–30, Dec. 2024, doi: 10.54554/jtec.2024.16.04.004.
- [7] L. Elvitaria *et al.*, "A Data-Driven Approach for Batik Pattern Classification Using Convolutional Neural Networks (CNN)," *Semarak International Journal of Electronic System Engineering*, 2024, doi: 10.37934/sijese.4.1.2230.

- [8] E. Sugiarto, A. Fahmi, and F. Budiman, "Implementation of Deep Learning Based on Convolution Neural Network for Batik Pattern Recognition," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 2025, doi: 10.22219/kinetik.v10i1.2019.
- [9] A. Y. Wicaksono, N. Suciati, C. Fatichah, K. Uchimura, and G. Koutaki, "Modified Convolutional Neural Network Architecture for Batik Motif Image Classification," *IPTEK Journal of Science*, vol. 2, no. 2, 2017, doi: 10.12962/j23378530.v2i2.a2846.
- [10] A. Hardirega and I. Jaelani, "Implementasi Convolutional Neural Network (CNN) Klasifikasi Motif Batik Menggunakan Efficientnet-B1," vol. 8, no. 5, pp. 10023–10028, 2024.
- [11] K. Azmi, S. Defit, and S. Sumijan, "Implementasi convolutional neural network (CNN) untuk klasifikasi batik tanah liat sumatera barat," *Jurnal Unitek*, vol. 16, no. 1, pp. 28–40, 2023.
- [12] K. B. Obaid, S. Zeebaree, O. M. Ahmed, and others, "Deep learning models based on image classification: a review," *International Journal of Science and Business*, vol. 4, no. 11, pp. 75–81, 2020.
- [13] Imam Fathurrahman, Mahpuz, Muhammad Djamaluddin, Lalu Kerta Wijaya, and Ida Wahidah, "Pengembangan Model Convolutional Neural Network (CNN) untuk Klasifikasi Penyakit Kulit Berbasis Citra Digital," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 1, pp. 298–308, Jan. 2025, doi: 10.29408/jit.v8i1.28655.
- [14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [15] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2013, pp. 8599–8603. doi: 10.1109/ICASSP.2013.6639344.
- [16] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *Proceedings of Machine Learning Research*, M. Meila and T. Zhang, Eds., in *Proceedings of Machine Learning Research*, vol. 139. PMLR, 2021, pp. 10347–10357.
- [17] I. Fathurrahman and I. Gunawan, "Pengenalan Citra Logo Kendaraan Menggunakan Metode Gray Level Co-Occurrence Matrix (GlcM) dan Jst-Backpropagation," *Infotek : Jurnal Informatika dan Teknologi*, vol. 1, no. 1, pp. 47–55, Jan. 2018, doi: 10.29408/jit.v1i1.894.
- [18] I. Fathurrahman and F. Fathurrahman, "Klasifikasi Penentuan Penerima Program Keluarga Harapan (PKH) Menggunakan Algoritma Support Vector Machine (Svm) Pada Kantor Dinas Sosial Lombok Timur," *Infotek : Jurnal Informatika dan Teknologi*, vol. 3, no. 1, pp. 27–31, Feb. 2020, doi: 10.29408/jit.v3i1.1802.
- [19] I. Fathurrahman, A. M. Nur, and F. Farhurrahman, "Identifikasi Kematangan Buah Mentimun Berbasis Citra Digital Menggunakan Jaringan Syaraf Tiruan Backpropagation," *Infotek: Jurnal Informatika dan Teknologi*, vol. 2, no. 1, pp. 27–33, Jan. 2019, doi: 10.29408/jit.v2i1.976.
- [20] R. C. Gonzalez and R. Woods, "Digital image processing: Pearson education india," *Digital image processing: Pearson education india*, 2009.
- [21] HendryHB, "Batik Nusantara (Batik Indonesia) Dataset," 2024, *Kaggle*. doi: 10.34740/KAGGLE/DSV/7641980.