

Analisis Perbandingan Akurasi Model EfficientNetB0 dan Vision Transformer Dalam Klasifikasi Citra Motif Batik Giriloyo

Ratna Puspita Sari, Albert Yakobus Chandra*

Fakultas Teknologi Informasi, Sistem Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia

Email: ¹211210053@student.mercubuana-yogya.ac.id, ²*albert.ch@mercubuana-yogya.ac.id

Email Penulis Korespondensi: albert.ch@mercubuana-yogya.ac.id

Submitted: 13/05/2025; Accepted: 31/05/2025; Published: 01/06/2025

Abstrak—Batik merupakan warisan budaya yang dimiliki oleh Indonesia dan telah diresmikan oleh UNESCO pada tanggal 2 Oktober 2009. Di era digital ini keragaman motif batik harus terus dilestarikan khususnya di Kampung Batik Giriloyo yang terletak di Karang Kulon, Wukirsari, Kec. Imogiri, Bantul. Kompleksitas dan keberagaman motif batik di daerah tersebut memerlukan pendekatan teknologi modern untuk membantu proses klasifikasi yang akurat. Penelitian ini bertujuan membandingkan kinerja kedua model terkini yaitu EfficientNetB0 dan *Vision Transformer* dalam mengklasifikasikan lima motif batik klasik di Kampung Batik Giriloyo. Metode penelitian ini menggabungkan pendekatan *deep learning* berbasis *Convolutional Neural Network (CNN)* dan *Transformer* dengan proses pelatihan dari nol tanpa adanya *transfer learning*. Tahapan penelitian yang digunakan meliputi pengumpulan dataset, *preprocessing*, augmentasi, bangun dan latih model, evaluasi serta visualisasi perbandingan hasil. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, F1-score serta efisiensi waktu inferensi. Dataset akhir berjumlah 13.128 citra batik hasil *slicing*. Dataset kemudian dibagi menjadi 3 bagian utama yaitu data pelatihan sebesar 80%, data validasi sebesar 10% dan data testing sebesar 10% dari total dataset. Hasil akhir penelitian menunjukkan bahwa *Vision Transformer* mencapai performa terbaik dengan akurasi pengujian mencapai 99.85% sedangkan model EfficientNetB0 memberikan hasil akurasi 98.78% dengan efisiensi yang stabil. Penelitian ini menegaskan bahwa model *Vision Transformer* lebih unggul dalam mengekstraksi pola global pada motif batik yang kompleks. Penelitian ini juga memberikan kontribusi nyata terhadap pemanfaatan kecerdasan buatan dalam pelestarian budaya melalui klasifikasi motif batik secara digital serta pengembangan sistem klasifikasi motif batik klasik di Kampung Batik Giriloyo.

Kata Kunci: Batik Giriloyo; *Deep Learning*; Klasifikasi Citra; EfficientNetB0; *Vision Transformer*

Abstract—Batik is a cultural heritage owned by Indonesia and has been inaugurated by UNESCO on October 2, 2009. In this digital era, the variety of batik motifs must be preserved, especially in Giriloyo Batik Village located in Karang Kulon, Wukirsari, Imogiri Sub-district, Bantul. The complexity and diversity of batik motifs in the area require a modern technological approach to assist the accurate classification process. This study aims to compare the performance of the two current models, EfficientNetB0 and Vision Transformer, in classifying five classic batik motifs in Kampung Batik Giriloyo. This research method combines deep learning approach based on Convolutional Neural Network (CNN) and Transformer with training process from zero without transfer learning. The research stages used include dataset collection, preprocessing, augmentation, model building and training, evaluation and visualization of result comparison. Evaluation is done using accuracy, precision, recall, F1-score and inference time efficiency metrics. The final dataset amounted to 13,128 sliced batik images. The dataset is then divided into 3 main parts, namely training data by 80%, validation data by 10% and testing data by 10% of the total dataset. The final results showed that Vision Transformer achieved the best performance with testing accuracy reaching 99.85 and the EfficientNetB0 model gave an accuracy of 98.78% with stable efficiency. This research confirms that the Vision Transformer model is superior in extracting global patterns in complex batik motifs. This research also makes a real contribution to the utilization of artificial intelligence in cultural preservation through the classification of digital batik motifs and the development of a classic batik motif classification system in Giriloyo Batik Village.

Keywords: Batik Giriloyo; Deep Learning; Image Classification; EfficientNetB0; Vision Transformer

1. PENDAHULUAN

Batik sebagai warisan dunia dan identitas budaya Indonesia yang sudah diresmikan pada tanggal 2 Oktober 2009 oleh UNESCO. UNESCO mengakui bahwa Batik Indonesia terdapat teknik serta simbol budaya yang tidak pernah hilang masanya. Keunikan seni tradisional ini didasari adanya lilin dan canting yang digunakan dalam menghasilkan pola ataupun corak yang rumit serta memiliki nilai estetika dan budaya yang tinggi. Setiap wilayah di Indonesia memiliki motif batik yang berbeda-beda sesuai filosofi atau ciri khas di wilayah tersebut [1]. Yogyakarta salah satu wilayah di Indonesia yang kental akan tradisi serta filosofi memberikan gambaran bahwa beragam motif batik jika dipelajari lebih lanjut. Kampung Batik Giriloyo salah satu lokasi di Yogyakarta yang sangat melestarikan batik klasik maupun batik cap tepatnya di Karang Kulon, Wukirsari, Kec. Imogiri, Bantul. Lokasi tersebut berdasarkan hasil wawancara merupakan sentra batik yang memiliki 12 paguyuban dan bekerja sama dimana tidak hanya melestarikan budaya tapi juga menawarkan berbagai macam jenis batik dan motifnya yang bisa dibeli langsung di galeri ataupun melalui pemesanan *online*. Keunikan lainnya dari Kampung Batik Giriloyo yaitu menyediakan paket wisata atau edukasi untuk masyarakat lokal maupun internasional dan bisa melestarikan langsung melalui proses membatik secara tradisional. Tentunya pelestarian batik di Kampung Batik Giriloyo menjadi sangat penting bagi generasi mendatang.

Era digital saat ini, teknologi pengenalan pola serta klasifikasi citra berbasis *deep learning* memberikan pengalaman untuk mekanisme dari identifikasi motif batik. Pendekatan ini mampu mengatasi keterbatasan metode manual yang membutuhkan waktu dan rentan terhadap kesalahan. Adanya model dengan kompleksitas yang lebih tinggi tentu bermanfaat dalam proses klasifikasi motif batik yang dilakukan secara akurat, cepat dan efektif [2].

Dua model *deep learning* modern yaitu EfficientNetB0 dan *Vision Transformer* (ViT) digunakan sebagai variabel utama dalam penelitian ini. EfficientNet itu sendiri merupakan salah satu model CNN pra-latih yang dikembangkan untuk meningkatkan performa pengenalan objek dengan efisiensi komputasi yang lebih baik [3]. Model EfficientNetB0 menggunakan pendekatan *compound scaling* untuk memaksimalkan *size*, jumlah filter serta kedalaman jaringan secara bersamaan. Pendekatan EfficientNetB0 memungkinkan model ini dalam mencapai akurasi tinggi dalam proses klasifikasi dengan jumlah parameter dan kebutuhan komputasi yang jauh lebih baik dibandingkan model CNN lainnya seperti ResNet, MobileNet, dan Inception. Kemampuannya dalam menyeimbangkan akurasi serta efisiensi, tentu banyak dibutuhkan dalam berbagai sistem pengenalan citra utamanya perangkat dengan keterbatasan sumber daya. Selain itu, *Vision Transformer* (ViT) adalah model kecerdasan buatan yang mengadaptasi model *transformer* yang sebelumnya sukses besar di pemrosesan bahan alami untuk tugas pengolahan citra. Proses dari ViT bisa dengan membagi ke patch kecil dan selanjutnya merepresentasikan sebagai *embedding* dan prosesnya menggunakan *self-attention*. Pendekatan ini memungkinkan ViT mampu menangkap relasi global setiap gambar secara lebih efektif berbeda halnya dengan CNN yang cenderung fokus pada fitur lokal melalui operasi konvolusi. Adanya kemampuan ini, ViT memiliki potensi lebih unggul dalam mengenali pola-pola kompleks dan variasi motif yang memiliki detail dan struktur geometris unik pada batik Kampung Giriloyo. Model EfficientNet dan *Vision Transformer* memiliki ciri khas arsitektur dan strategi ekstraksi fitur yang berbeda. Efektivitas model dan efisiensi dalam proses klasifikasi motif batik di Kampung Giriloyo menjadi fokus utama penelitian ini. Motif batik di Kampung Batik Giriloyo memiliki keunikan pada pola, warna serta detail aksesoris yang memerlukan kemampuan ekstraksi fitur dan generalisasi yang kuat dan baik. Perbandingan kedua model pada dataset motif batik klasik di Kampung Batik Giriloyo dalam konteks klasifikasi diharapkan mampu memberikan pengetahuan terkait keunggulan dan keterbatasan masing-masing arsitektur dan untuk pengembangan sistem pelestarian digital batik di masa mendatang mampu memberikan dasar pemilihan model yang terbaik. Evaluasi dilakukan menggunakan metrik akurasi serta efisiensi komputasi seperti waktu inferensi dan penggunaan memori atau jumlah parameter model yang menjadi aspek penting dalam implementasi sistem sistem klasifikasi citra batik pada perangkat yang memiliki keterbatasan sumber daya.

Penelitian ini meskipun menerapkan model *deep learning* yang telah berhasil diimplementasikan dalam klasifikasi citra khususnya pengenalan motif batik tentu memiliki kekurangan signifikan dalam membandingkan akurasi EfficientNetB0 dan *Vision Transformer* (ViT) pada dataset motif batik Kampung Giriloyo. Karakteristik pola yang kompleks serta detail yang khas sehingga menuntut model untuk dapat menangkap fitur lokal dan global secara efektif. Selain itu, tantangan yang dihadapi dalam penerapan teknologi ini dalam keterbatasan data dan sumber daya komputasi di lingkungan pengrajin batik. Maka dari itu, muncul pembahasan terkait model mana yang lebih unggul dalam hal akurasi dan efisiensi komputasi dalam klasifikasi motif batik di Kampung Batik Giriloyo secara otomatis [4].

Pembahasan beberapa tahun terakhir tentang penggunaan *deep learning* untuk klasifikasi motif batik sudah mengalami peningkatan. EfficientNet yang termasuk bagian model *Convolutional Neural Network* (CNN) telah banyak diimplementasikan karena mampu mengekstraksi fitur secara otomatis dan menghasilkan akurasi yang tinggi. Dalam penelitian yang berjudul “*Deep Learning Untuk Klasifikasi Motif Batik Papua Menggunakan EfficientNet dan Transfer Learning*” di publikasikan pada tahun 2022 membahas tentang pengaplikasian EfficientNetB2 melalui teknik *transfer learning* dan augmentasi data dalam klasifikasi motif batik Papua kemudian mampu meningkatkan akurasi menjadi 90% dari yang semula hanya 72% setelah adanya tahapan augmentasi seperti *Color Jitter dan Contrast* [2]. Tentu penelitian ini menunjukkan potensi terbaik EfficientNet dalam mengidentifikasi pola batik yang kompleks dan detail dengan efisiensi komputasi yang baik. Studi lainnya pada tahun 2024 juga menyampaikan bahwa EfficientNetV2 mampu memberikan performa klasifikasi terbaik dengan akurasi validasi mendapatkan hasil 97,6% dan akurasi pengujian sebesar 96% pada dataset batik yang beragam mengungguli beberapa arsitektur populer seperti Xception, ResNet50V2 dan MobileNetV2 ditulis dalam jurnal yang berjudul “*Hybrid Deep Learning EfficientNetV2 and Vision Transformer for Multi-Class Image Classification*” [5]. Selain itu penelitian tentang klasifikasi batik lampung juga menunjukkan keunggulan EfficientNet dibandingkan AlexNet, LeNet dan MobileNet dalam mengidentifikasi 10 kelas motif batik lokal dengan hasil yang kompetitif [1].

Dalam konteks lain, *Vision Transformer* telah berevolusi pengenalan citra mekanisme *self-attention* yang efektif dalam mengambil integritas antar fitur gambar sehingga dapat mengetahui pola yang kompleks dan variatif. Studi internasional menunjukkan bahwa terdapat keunggulan di masing-masing model yaitu EfficientNet dan *Vision Transformer* yang tergantung pada karakteristik dataset dan kompleksitas pola atau corak yang dianalisis [6]. Akan tetapi, masih sedikit yang memanfaatkan dan menilai kedua model ini pada motif batik Indonesia khususnya motif di Kampung Batik Giriloyo yang sangat filosofis dan detail. Gap penelitian yang muncul minim akan studi komparatif antara EfficientNet dan *Vision Transformer* dalam klasifikasi motif batik tradisional Indonesia dimana kurang akan eksplorasi pada motif lokal yang kompleks seperti di Kampung Batik Giriloyo dan kurang evaluasi aspek efisiensi komputasi yang penting untuk implementasi di lingkungan dengan keterbatasan perangkat keras. Maka dari itu, penelitian ini digunakan untuk mengisi kekurangan tersebut dengan melakukan analisis akurasi antara EfficientNetB0 dan *Vision Transformer* dalam klasifikasi motif batik yang nantinya mampu memberikan pilihan model terbaik untuk melestarikan budaya melalui digitalisasi.

Penelitian ini bertujuan untuk melakukan perbandingan dan analisis akurasi antara model EfficientNetB0 dan *Vision Transformer* dalam klasifikasi motif batik dengan studi kasus Kampung Batik Giriloyo. Fokus pada aspek penting yaitu klasifikasi serta efisiensi komputasi. Kedua model dipilih berdasarkan keunggulan yang telah dibuktikan

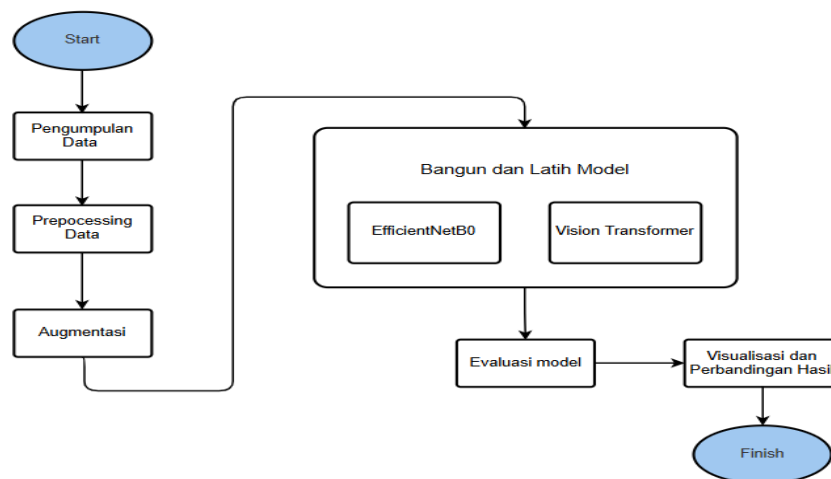
dalam berbagai penelitian atau studi sebelumnya. Salah satu penelitian yang menunjukkan bahwa EfficientNetB0 dikenal mampu memberikan efisiensi parameter dan kecepatan pelatihan yang tinggi dan stabil tanpa mengurangi peningkatan akurasi, sedangkan *Vision Transformer* menawarkan kemampuan unggul dalam mekanisme *self-attention* walaupun perlu sumber daya komputasi yang besar. Pendekatan ini diharapkan dapat memberikan dampak jelas terkait model yang terbaik untuk digunakan dalam klasifikasi motif batik baik dari akurasi maupun efisiensi sehingga mampu menjadi dasar bagi pengembangan sistem pelestarian budaya dengan teknologi kecerdasan buatan yang efektif dan efisien [7].

Penelitian ini juga mengangkat studi kasus di Kampung Batik Giriloyo karena sangat relevan dengan konsep motif batik klasik yang masih tradisional. Pelestarian motif batik di daerah tersebut memerlukan dukungan teknologi yang mendukung proses pengenalan dan klasifikasi secara akurat dan efisien. Seiring perkembangan industri kreatif dan digitalisasi budaya, penerapan model *deep learning* yaitu EfficientNetB0 dan *Vision Transformer* menjadi sangat penting dalam menunjang transformasi digital bidang kerajinan batik. Hal ini sejalan dengan kebutuhan untuk mengembangkan serta meningkatkan daya saing produk batik lokal di pasar nasional maupun internasional melalui inovasi teknologi yang mampu memproses dengan cepat identifikasi motif serta dokumentasi data motif batik secara terstruktur [8]. Penelitian ini dengan demikian akan memiliki urgensi tinggi untuk memberikan kontribusi nyata dalam pengembangan dan pelestarian teknologi digital dalam klasifikasi motif batik yang efisien melalui pemanfaatan kecerdasan buatan. Hasil penelitian diharapkan mampu menjadi dasar bagi pengembangan sistem berbasis *deep learning* yang dapat diakses oleh pengrajin dan pelaku industri serta mempertahankan posisi Kampung Batik Giriloyo sebagai pusat budaya dan inovasi digital.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini mengimplementasikan pendekatan kuantitatif eksperimental yang mengacu pada prinsip-prinsip data mining dan klasifikasi yang banyak digunakan dalam penelitian komputer dan dilakukan melalui tahapan yang sistematis dan terstruktur [9]. Tahapan penelitian ini dirancang untuk menganalisis dan membandingkan akurasi dua model *deep learning* yaitu model EfficientNetB0 dan *Vision Transformer*. Adanya penelitian ini sesuai dengan metode yang digunakan dalam penelitian klasifikasi citra berbasis *deep learning* terdahulu. Dalam klasifikasi citra motif batik di Kampung Batik Giriloyo terdapat tahapan-tahapan sistematis yang dilakukan seperti pada Gambar 1 berikut :



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Pengumpulan data termasuk tahap awal dalam melakukan sebuah penelitian baik pengumpulan data gambar, angka ataupun lainnya yang relevan dengan kondisi penelitian. Dataset yang digunakan merupakan kumpulan citra motif batik Kampung Giriloyo yang didapatkan dengan mengumpulkan data dari beberapa sumber yaitu dokumentasi digital dari arsip di Galeri Kampung Batik Giriloyo dan pengambilan gambar secara langsung dengan perangkat seperti kamera beresolusi tinggi. Pengumpulan data ini bertujuan untuk memastikan dataset yang komprehensif dan representatif serta relevan dengan pendekatan yang digunakan dalam penelitian klasifikasi citra lainnya [10].

2.3 Preprocessing

Proses *preprocessing* dan augmentasi dilakukan setelah tahapan *slicing* dan *splitting* guna memastikan data konsisten dan menghindari kebocoran data. Detail proses yang ada pada tahapan penelitian ini yaitu *resizing* gambar, *slicing* citra, dan pembagian dataset atau *splitting* [11].

2.3.1 Slicing Citra

Teknik ini digunakan untuk mendapatkan data yang variatif dan model dapat mempelajari detail motif batik klasik di Kampung Batik Giriloyo. Tiap citra motif batik dipotong menjadi 4 bagian yang sama, dengan begitu mampu meningkatkan sampel data pelatihan tanpa melakukan pengumpulan data tambahan. Dalam penelitian lainnya yaitu pemeriksaan CT-Scan hal yang mempengaruhi kualitas citra yang dihasilkan adanya variasi *slice thickness* sehingga konteks ini dapat digunakan sebagai gambaran bagaimana teknik ini menunjukkan sudut pandang dan detail berbeda pada model.

2.3.2 Pembagian Dataset (*Splitting*)

Setelah tahapan *slicing* dilakukan, selanjutnya yaitu dengan membagi data menjadi 3 bagian utama yaitu data pelatihan sebesar 80% , data validasi sebesar 10% dan data testing sebesar 10% dari total dataset [6]. Data pelatihan digunakan untuk melatih model , data validasi untuk memantau kinerja model selama proses pelatihan dan melakukan penyesuaian parameter, dan data pengujian untuk menguji kinerja akhir model setelah pelatihan selesai.

2.4 Augmentasi Data

Tahapan ini dilakukan untuk membantu ketersediaan dataset dengan menambahkan jumlah dataset untuk dilakukan manipulasi data seperti *rescale*, rotasi, *zoom*, horizontal dan lainnya. Augmentasi juga termasuk tahapan peningkatan ukuran, jumlah serta kualitas yang dapat digunakan untuk mengatasi permasalahan keterbatasan data yang dapat menyebabkan *overfitting*.

2.5 Bangun dan Latih Model

Model *deep learning* yang digunakan terdapat 2 fokus yaitu EfficientNetB0 dan *Vision Transformer*. Pada tahapan ini model dipilih berdasarkan karakteristik dan keunggulan tiap pengolahan citra. Tahap ini akan dilakukan pembangunan dan pelatihan model dari nol menggunakan dataset motif batik tanpa menggunakan proses *transfer learning*. Tahapan pelatihan model dilakukan secara menyeluruh dari awal menggunakan dataset yang telah di preprocessing dan augmentasi. Parameter yang dimanfaatkan dalam pelatihan adalah *optimizer*, *batch size*, *learning rate*, *loss function*, *metric*, *callbacks*, augmentasi dan epoch. Kinerja model pada masa pelatihan dilihat menggunakan data validasi sebanyak 10% dari dataset untuk memastikan model mempelajari data dengan baik. Pelatihan model dari nol bertujuan menguji kemampuan EfficientNetB0 dan *Vision Transformer* dalam mengetahui motif batik di Kampung Batik Giriloyo tanpa terpengaruh oleh *bobot pretrained* ataupun *transfer learning*. Latih model dilakukan menggunakan parameter diantaranya *batch size* dan epoch dengan mekanisme *early stopping* jika akurasi validasi tidak meningkat maka dapat dilakukan penambahan epoch. Selain itu, pelatihan menggunakan CPU intel i5 dengan RAM 16 tanpa GPU yang menyesuaikan kondisi sumber daya komputasi. Implementasi CPU sebagai perangkat pelatihan juga menguji kelayakan implementasi model di lingkungan dengan keterbatasan infrastruktur komputasi.

2.5.1 EfficientnetB0

EfficientnetB0 merupakan varian dari model EfficientNet yang dibangun dari awal (*scratch*) tanpa *transfer learning* agar dapat melakukan perbandingan secara terstruktur. Model dipilih dengan mempertimbangkan efisiensi penggunaan sumber daya komputasi dan keunggulan dalam mencapai akurasi terbaik dengan parameter yang relatif sedikit. EfficientNetB0 menggunakan konsep skala model yang dipelajari secara sistematis dalam menyeimbangkan lebar, kedalaman, dan resolusi jaringan sehingga didapatkan hasil model yang adaptif dan efisien [12]. Model ini dibangun dengan parameter awal dan dilatih sepenuhnya menggunakan dataset yang telah dikumpulkan sehingga proses pelatihan bertujuan menemukan representasi fitur yang optimal untuk motif batik yang kompleks dan detail. Konsep dasar untuk adaptasi model yang dibangun yaitu penggunaan aktivitas input lapisan dengan input gambar (224, 224, 3). Pelatihan data pada model EfficientNetB0 menggunakan pelatihan dari nol tanpa adanya *transfer learning*. Model yang dibangun mengadaptasi konsep dasar EfficientNet berupa penggunaan aktivasi *Swish*, *Blox*, *MBCConv*, *Depthwise Separable Convolutional* untuk efisiensi serta penambahan *residual connection* untuk mencegah *vanishing gradient*. Tahapan model ini menggunakan layer pertama yaitu Conv2D dengan 32 filter dan kernel 3x3 diikuti BatchNormalization dan MaxPooling2D. Setelah proses ekstraksi fitur, digunakan GlobalAveragePooling2D dan Dropout serta keluaran Deanse layer dengan softmax sebanyak 5 neuron.

2.5.2 Vision Transformer

Vision Transformer termasuk arsitektur yang berbeda secara fundamental dari *Convolutional Neural Network* (CNN) karena menggunakan mekanisme *self-attention* untuk memproses citra sebagai urutan dari patch-patch. Pemilihan model ini tentu berpotensi memberikan hasil lebih baik terutama dataset kompleks dengan variasi pola yang tinggi dan rumit seperti motif batik dan model harus mengenali pola dan hubungan antar patch secara mandiri dari dataset batik di Kampung Batik Giriloyo [13].

2.6 Evaluasi Model

Metrik evaluasi yang digunakan untuk mengevaluasi kinerja model telah banyak direkomendasikan dalam penelitian klasifikasi citra yaitu terdiri dari akurasi, presisi, recall, F1- score dan efisiensi komputasi. Lima metrik ini penting

untuk menilai potensial model dalam mengklasifikasikan motif batik secara menyeluruh, serta efisiensi dalam penggunaan sumber daya komputasi. Akurasi (*Accuracy*) digunakan untuk mengukur proporsi prediksi yang relevan dari semua data pengujian. Metrik ini mempertimbangkan jumlah prediksi benar dari kelas positif dan negatif. Secara sistematis, berikut akurasi yang telah dirumuskan:

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

Penjelasan rumus akurasi menunjukkan dimana TP (*True Positive*) merupakan jumlah prediksi benar untuk kelas positif, TN (*True Negative*) menunjukkan jumlah prediksi benar untuk kelas negatif, kemudian FP (*False Positive*) adalah jumlah prediksi salah untuk kelas negatif yang dikategorikan sebagai positif dan FN (*False Negative*) mengacu pada jumlah prediksi salah untuk kelas positif yang diklasifikasikan sebagai negatif. Presisi (*Precision*) dan Recall merupakan metrik evaluasi ini digunakan untuk mengamati evaluasi performa model pada tiap kelas motif batik terkait keseimbangan dataset. Presisi mengukur rasio prediksi positif yang sesuai terhadap seluruh prediksi positif, sebagaimana ditunjukkan dalam rumus berikut:

$$Presicion = \frac{TP}{TP+FP} \quad (2)$$

Rumus presisi menjadi penting ketika kesalahan klasifikasi positif (*false positive*) mempunyai konsekuensi besar seperti dalam klasifikasi penyakit atau data motif yang mirip secara visual. Kemudian pada rumus recall digunakan untuk mengukur kemampuan model dalam mengidentifikasi seluruh data kelas positif atau yang disebut juga *sensitivity* atau *true positive rate*. Metrik ini sangat penting apabila kegagalan mendeteksi data positif merupakan kesalahan yang signifikan. Selanjutnya recall menunjukkan seberapa banyak dari total kelas positif yang berhasil dikenali dengan benar oleh model seperti ditunjukkan pada rumus berikut ini:

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

F1-Score merupakan matrik evaluasi yang bertujuan untuk memberikan gambaran keseimbangan antara matrik presisi dan recall. F1-Score menggabungkan presisi dan recall pada satu nilai harmonik rata-rata. Rumus F1-Score dituliskan sebagai berikut:

$$F1 = 2 \times \frac{Presicion \times Recall}{Presicion+Recall} \quad (4)$$

Pada rumus ini menunjukkan semakin tinggi nilai F1-Score maka akan semakin baik kinerja model dalam menyeimbangkan antara prediksi benar dan kegagalan mendeteksi kelas yang seharusnya.

Efisiensi komputasi mengukur waktu inferensi (*inference time*) dan waktu pelatihan (*training time*) tiap gambar dan penggunaan memori selama proses pengujian berlangsung untuk menilai kelayakan model dalam implementasi nyata. Efisiensi komputasi mencakup beberapa aspek untuk menentukan kelayakan model saat diterapkan nyata seperti sistem berbasis edge computing atau pada perangkat dalam kemampuan terbatas.

2.7 Visualisasi dan perbandingan hasil

Hasil pelatihan dan evaluasi model dirangkum dalam sebuah visualisasi yang mudah dipahami serta mampu memberikan gambaran jelas akan performa model yang sedang dianalisis. Adapun visualisasi yang digunakan dalam penelitian ini meliputi grafik akurasi dan *loss* per epoch, *confusion matrix*, tabel metrik evaluasi serta *bar chart* metrik. Pada visualisasi grafik digunakan untuk menunjukkan peningkatan akurasi dan nilai *loss* selama tahap pelatihan yang membantu dalam mengenali model apakah mengalami *overfitting* atau *underfitting*. Visualisasi *confusion matrix* dimanfaatkan untuk menampilkan distribusi prediksi tiap kelas motif batik yang sulit dikenali. Kemudian pada tabel metrik evaluasi menyajikan perbandingan nilai *accuracy*, *presicion*, *recall* serta *F1-Score* yang telah diproses sebelumnya di tiap model secara kuantitatif. Pada *bar chart* per metrik menunjukkan perbandingan metrik antar model dalam bentuk diagram batang sebagai visualisasi untuk menginterpretasi performa relatif masing-masing model.

3. HASIL DAN PEMBAHASAN

Pada bagian membahas hasil penelitian yang telah dilakukan dan mengulas analisis perbandingan akurasi model. Fokus penelitian ini pada klasifikasi motif batik di Kampung Batik Giriloyo dengan menerapkan metode *Convolutional Neural Network* (CNN) dengan model EfficientNetB0 yang dibandingkan dengan *Vision Transformer* (ViT). Bagian ini juga memaparkan hasil penelitian secara objektif berdasarkan data yang didapatkan dari pengkajian klasifikasi motif batik Kampung Giriloyo menggunakan dua model *deep learning* tersebut. Berdasarkan analisis terhadap penelitian terdahulu, penerapan metode ini masih jarang dibandingkan dan dapat menjadi solusi yang tepat untuk mengklasifikasikan motif batik di Kampung Batik Giriloyo secara lebih akurat dan sistematis. Hasil ditampilkan dalam bentuk metrik evaluasi, tabel, grafik dan visualisasi yang menggambarkan komprehensif mengenai performa masing-masing model. Penilaian performa model juga menjadi indikator utama dengan akurasi sebagai alat ukur tingkat relevansi prediksi model pada label sebenarnya. Adapun hasil evaluasi model memberikan analisis mendalam

terhadap kelebihan dan kekurangan masing-masing model serta kapabilitas dalam menangani kelas yang memiliki pola rumit. Berikut penjelasan dari hasil tahapan penelitian yang sudah dilakukan [14].

3.1 Hasil Pengumpulan Dataset

Hasil Dataset ini mewakili kelas motif batik yang digunakan dalam penelitian yaitu mencakup 5 kelas motif batik dimana berasal dari Galeri Kampung Batik Giriloyo. Motif batik yang diambil termasuk jenis motif batik klasik yaitu Batik Kawung, Batik Sidoasih, Batik Sidomukti, Batik Truntum, dan Batik Wahyu Tumurun. Dataset yang terkumpul yaitu 3538 dimana gambar batik kawung berjumlah 514 gambar, batik sidoasih 777 gambar, batik sidomukti 692 gambar, batik truntum 1036 gambar dan batik wahyu tumurun berjumlah 519 gambar. Setelah dilakukannya tahap awal dan didapatkan data yang akan digunakan kemudian dikelompokkan berdasarkan direktori masing-masing agar lebih terstruktur dan sistematis untuk siap untuk menghasilkan direktori lainnya yang relevan dengan tahap selanjutnya. Berikut merupakan hasil pengumpulan dataset yang dikelompokkan berdasarkan kategori atau kelasnya pada Gambar 2 seperti dibawah ini.



Gambar 2. Contoh Data Citra Motif Batik Giriloyo

Gambar 2 diatas menunjukkan masing-masing satu contoh dari data lima kelas motif motif yang digunakan dalam penelitian. Pada gambar (a) yaitu Motif Batik Kawung mempunyai pola geometris elips yang tersusun rapi seperti kolang-kaling (buah aren) dimana melambangkan pengendalian diri dan kesucian. Gambar (b) Motif Batik Sidoasih terdiri dari motif ornamen bunga dan susunan simetris yang menggambarkan harapan akan kehidupan yang penuh kasih sayang dan cinta yang tulus dimana motif ini berasal dari gabungan kata “sido” (menjadi) dan “asih” (kasih sayang). Ditunjukkan pada gambar (c) yaitu Motif Batik Sidomukti juga memiliki makna dari kata “sido” (menjadi) dan “mukti” (bahagia dan makmur) yang menjadi simbol kemuliaan dan kemakmuran. Kemudian gambar (d) dimana Motif Batik Truntum tercipta dari kisah cinta abadi raja Pakubuwono III dan permaisurinya dimana truntum berasal dari kata “tumaruntum” (bersemi kembali) dan biasa digunakan oleh orang tua pengantin sebagai simbol cinta tanpa syarat dan bimbingan orang tua. Pada gambar (e) merupakan gambaran visual dari Motif Batik Wahyu Tumurun ditunjukkan simbol seperti huruf W dan mempunyai arti yang menggambarkan harapan akan datangnya wahyu atau berkah dari Tuhan, Ornamen Mahkota, burung garuda dan sayap yang menjadi lambang kedudukan tinggi dan karunia ilahi dan digunakan sebagai simbol doa untuk kesuksesan, jabatan dan keberuntungan.

3.2 Preprocessing

Pengambilan Dataset di awal menghasilkan dataset asli yang berjumlah total 3.538 gambar yang disimpan pada folder dataset skripsi. Setelah tahapan pengambilan dataset selesai, perlu adanya tahapan *preprocessing* dimana terbagi atas *slicing citra* dan *splitting*. Hasil tahap ini bertujuan untuk mempersiapkan data untuk dapat digunakan secara optimal dalam bangun dan latih model. Selain itu, *preprocessing* yang dilakukan secara sistematis juga dapat meningkatkan kualitas data yang berpengaruh terhadap tingkat akurasi dan efektivitas model dalam menghasilkan analisis yang valid dan mendalam.

3.2.1 Slicing Citra

Data dalam konteks *computer vision* dan *machine learning* merujuk pada proses pembagian data, rekaman atau citra menjadi bagian-bagian subset tertentu. Hasil dari tahapan *slicing* ini diharapkan mampu meningkatkan variasi data dengan membaginya menjadi 4 potongan dengan melewati gambar rusak dan jumlah data yang tersedia untuk pelatihan model bertambah maka model mampu belajar lebih banyak variasi dan detail. Bagian yang dipotong meliputi *top left*, *top right*, *bottom left*, dan *bottom right*. Selain itu adanya *slicing* gambar pada dataset yang cukup terbatas memudahkan analisis subset data yang berguna untuk evaluasi performa model lebih mendalam. Melakukan *slicing* juga mampu memperbaiki kinerja model pada data tertentu dan mempercepat serta menghemat proses pelatihan sehingga menjadi lebih efisien [15].

Tabel 1. Dataset Jumlah Gambar Sebelum dan Sesudah Proses *Slicing*

Kelas	Jumlah Gambar	
	Sebelum <i>Slicing</i>	Setelah <i>Slicing</i>
Batik_kawung	514	2056
Batik_sidoasih	777	3108
Batik_sidomukti	692	2748
Batik_truntum	1036	4076
Batik_wahyutumurun	519	1140

Seperti pada Tabel 1 diatas, proses *slicing* dilakukan dengan membagi setiap citra dengan konfigurasi yang sudah ditentukan yaitu *image size* sebesar 224 x 224 dengan menyesuaikan kebutuhan model, *batch image* berukuran 16 yang mengikuti batch kecil untuk kondisi perangkat. Penerapan *slicing* juga membantu mengurangi resiko *overfitting* dan meningkatkan generalisasi model. Selain itu, proses ini juga membantu dalam memproses dataset yang rusak agar dilewati sehingga tidak mengganggu tahapan kedepannya. Jumlah dataset juga mempengaruhi berjalannya model dimana tanpa adanya *slicing* atau dengan jumlah terbatas hasil akurasi model EfficientNetB0 hanya mendapatkan akurasi yang rendah dan nilai *loss* yang tidak signifikan.

3.2.2 Splitting Data

Hasil dari tahapan ini apabila telah melalui proses *slicing* kemudian disimpan dalam direktori baru yang telah dibagi menjadi 3 bagian yaitu pelatihan (80%), validasi (10%), dan test (10%). Tahapan ini dilakukan untuk mencegah bias model pada kelas tertentu dan meningkatkan generalisasi model. Pembagian ini dilakukan dengan teknik pengambilan sampel menjadi kelompok kecil berdasarkan karakteristik tertentu yang relevan atau *stratified* kelas. Terdapat penelitian yang membahas *stratified* membantu melindungi representativitas data pada setiap subset untuk pelatihan model yang adil dan akurat [16]. Proses pembagian data menggunakan skrip Python yang memanfaatkan modul *os*, *shutil* dan *random* yang secara otomatis dapat membagi gambar secara acak sesuai rasio yang ditentukan. Tujuan dari tahapan ini yaitu digunakan untuk memaksimalkan jumlah data pelatihan agar model dapat mengidentifikasi pola yang variatif, menyediakan data validasi untuk monitoring *overfitting*, dan menyediakan data pengujian baru bagi model untuk menilai performa generalisasi. Pentingnya komposisi pembagian data sangat mempengaruhi performa akhir model. Apabila data pelatihan terlalu kecil dalam hal proporsi maka model akan kekurangan informasi untuk dipelajari. Begitupun jika data pengujian terlalu sedikit maka evaluasi performa model akan mengalami kurang representatif dan rawan bias. Penelitian terbaru menjelaskan bahwa estimasi performa model akurat dan terhindar dari kebocoran informasi [17].

3.3 Augmentasi

Tahap augmentasi penting dalam pelatihan model untuk mengoptimalkan jumlah data yang terbatas dan kasus klasifikasi motif batik yang memiliki detail cukup rumit. Selain itu, digunakannya augmentasi untuk mendapatkan hasil generalisasi model yang baik serta mengurangi *overfitting*. Pemanfaatan augmentasi berupa *rescale*, *rotasi*, *zoom*, serta *horizontal flip* tentu membantu proses ini dalam peningkatan jumlah serta memperluas keragaman data tanpa menambahkan data baru secara manual [18]. Dalam penelitian ini, penerapan augmentasi dilakukan setelah tahap *preprocessing* selesai untuk dapat mengenali lebih dalam pola batik klasik Kampung Giriloyo. Gambar 3 menunjukkan proses augmentasi pada salah satu motif batik yaitu batik Kawung. Teknik augmentasi ini bertujuan untuk meningkatkan variasi data pelatihan dan memperbaiki generalisasi model.



Gambar 3. Proses Augmentasi Data pada Motif Batik Kawung

Berdasarkan Gambar 3, citra sebelah kiri atau original merupakan citra asli sebelum dilakukan augmentasi kemudian sebelah citra sebelah kanan yaitu Aug-1 sampai Aug-5 merupakan hasil dari beberapa teknik augmentasi. Pada Aug-1 hasil data dengan rotasi acak dan *zoom* untuk menunjukkan gambar dari sudut yang berbeda. Aug-2 menampilkan hasil pergeseran posisi pola untuk menambahkan variatif objek. Kemungkinan model dalam mengenali motif dengan berbagai posisi seperti miring juga ditampilkan pada hasil Aug-3. Pada Aug-4 dan Aug-5 dimana *flip horizontal* dan *brightness adjustment* juga diterapkan untuk menghindari ketergantungan pada arah pola dan pencahayaan berbeda pada pengambilan gambar.

3.4 Hasil Pelatihan Model

3.4.1 EfficientNetB0

Pada bagian ini, proses pelatihan *Convolutional Neural Network* (CNN) dengan model EfficientNet tanpa *transfer learning* untuk klasifikasi lima kelas motif batik dari Kampung Batik Giriloyo. Model yang dilatih menggunakan 100 epoch, dan penambahan *Earlystopping* dengan *patience* yang digunakan untuk menghentikan otomatis proses pelatihan lebih awal saat model tidak mengalami peningkatan performa data validasi setelah jumlah epoch tertentu serta menghindari *overfitting* dan menghemat waktu komputasi. Berdasarkan proses yang sudah dijalankan, pelatihan model ini dihentikan lebih awal pada epoch 73 karena menampilkan performa yang sangat baik dan stabil. Proses ini menunjukkan bagaimana model belajar secara bertahap dari data citra dan bagaimana performa mengalami

peningkatan. Model EfficientNetB0 memberikan performa yang impresif dalam klasifikasi citra motif batik meskipun dilatih tanpa bobot awal dari *pretraining*.

Pada epoch awal, model mendapatkan hasil akurasi senilai 53,62% dengan akurasi validasi 59,39%. Nilai ini menjadi titik awal yang baik karena bobot model masih acak dan belum memiliki pemahaman apapun terkait karakteristik citra motif batik. Beberapa epoch awal juga menunjukkan adanya peningkatan secara signifikan yaitu epoch kedua menghasilkan 75,19% dan epoch ketiga 80,65% secara urut. Hasil ini menunjukkan bahwa model dapat dengan cepat menangkap fitur visual dasar yang kompleks seperti garis, warna serta pola dari motif batik. Selama pelatihan lima epoch pertama mengalami penurunan secara bertahap dari 1.2227 ke 0,04081. Sementara itu, *loss* pada data validasi juga menurun menggambarkan bahwa model belajar dengan baik dan dapat meningkatkan kemampuan generalisasi. Nilai akurasi validasi pada epoch kelima senilai 79,47% dan bisa disimpulkan bahwa model EfficientNetB0 berhasil mengenali pola penting pada citra dan mampu membedakan adanya kemiripan visual di tiap kelas motif batik. Memasuki pertengahan pelatihan model masih menunjukkan peningkatan yang signifikan dengan akurasi mencapai lebih dari 90% yaitu pada epoch ketujuh dan delapan namun terdapat fluktuasi dengan turunnya akurasi validasi pada epoch kesembilan dimana senilai 66,72% dengan validasi *loss* mencapai 1.3683. Hal ini memungkinkan terjadinya *overfitting* terhadap batch tertentu selama proses validasi. Karena pelatihan model dari nol maka model lebih rentan terjadi perubahan pada beberapa epoch. Namun hal tersebut bersifat sementara, memasuki epoch ke sebelas dan selanjutnya akurasi validasi kembali meningkat dengan hasil 92,44% dan 82,44%. Model masih dapat kembali beradaptasi dan menyempurnakan bobot dengan baik.

Setelah fase fluktuasi dilalui, model masih dapat mempertahankan akurasi pelatihan melebihi 95% dan akurasi validasi melebihi 90%. Nilai *loss* pada pelatihan maupun validasi juga menunjukkan penurunan yang signifikan dan cenderung stabil menandakan model dapat mencapai pemahaman kuat akan pola penting dari tiap kelas motif batik. Pada epoch 73 menunjukkan hasil akurasi pelatihan tertinggi mencapai 98,82% dan akurasi validasi 98,09%. Nilai *loss* validasi juga sangat rendah senilai 0,0587. Performa model optimal dalam mengidentifikasi serta mengklasifikasikan motif batik pada presisi yang tinggi juga. Berdasarkan kurva pelatihan dan validasi, pada tahap akhir tidak terjadi gejala *overfitting* yang signifikan. Maka dari itu kombinasi struktur EfficientNet dan *preprocessing* yang tepat selain menghasilkan model yang akurat tetapi juga generalisasi terhadap data baru.

3.4.2 Vision Transformer (ViT)

Seperti model EfficientNetB0, hasil dan analisis dari proses pelatihan model *Vision Transformer* berdasarkan nilai akurasi dan *loss* pada data pelatihan maupun validasi. Proses pelatihan juga dilakukan selama 100 epoch dengan *earlystopping* di epoch 92 karena sebelum mencapai target model telah mencapai konvergensi dan *loss* validasi sangat rendah <0.01. Seperti model EfficientNetB0, hasil dan analisis dari proses pelatihan model ini berdasarkan nilai akurasi dan *loss* yang dihasilkan pada data pelatihan maupun validasi. Tahap awal pelatihan model *Vision Transformer* pada epoch pertama menunjukkan akurasi pelatihan senilai 39,42% dan akurasi validasi senilai 66,11%. Meskipun performa pelatihan masih rendah namun model mulai mengenali pola dasar dari data validasi. Pada epoch kedua terdapat kenaikan akurasi pelatihan dan validasi cukup signifikan menjadi 68% dan akurasi validasi mencapai 87,86%. Penurunan nilai *loss* validasi dari 0.8898 menjadi 0.4255 membuktikan bahwa model mulai menangkap pola penting pada citra batik. Kenaikan akurasi ini terus berlanjut pada epoch ketiga hingga kelima dengan akurasi pelatihan senilai 90,87% dan akurasi validasi tetap tinggi meskipun terjadi fluktuatif menandakan proses pelatihan yang stabil. Menuju epoch keenam sampai sepuluh menunjukkan Tingkat stabilitas dan efektivitas dalam proses belajar. Epoch ketujuh menunjukkan akurasi validasi 94,05% dengan *loss* validasi 0.1994. Menuju pertengahan ini, model *Vision Transformer* menunjukkan performa yang sangat tinggi mencapai 98,61% dan akurasi validasi maksimal 99,24% serta nilai *loss* 0.028 dimana tingkat kesalahan klasifikasi sangat kecil. Naik turunnya akurasi bisa disebabkan oleh *noise* pada data validasi atau variasi kecil dalam struktur motif yang membuat perlu kesesuaian ulang model. Pada akhir pelatihan kekuatan akurasi validasi mencapai titik tertinggi 99,69% dengan nilai *loss* validasi terendah 0.0061 yang terjadi pada epoch 49. Capaian membuktikan kekuatan *Vision Transformer* dapat mempelajari fitur visual secara mendalam dan generalisasi yang sangat baik pada data validasi yang belum pernah ditampilkan sebelumnya.

3.5 Evaluasi Model

Setelah melakukan pelatihan model, perlu adanya evaluasi kinerja pada data pengujian. Evaluasi dilakukan untuk menghasilkan kemungkinan dari kekurangan pada citra yang dibaca pada proses klasifikasi, jadi akan menghasilkan nilai akurasi dan nilai *loss* dengan peluang tertinggi yang akan didapatkan dari semua model pengujian. Penelitian ini bertujuan untuk membandingkan akurasi model EfficientNetB0 dan *Vision Transformer* dalam klasifikasi citra motif batik Kampung Giriloyo. Evaluasi yang digunakan dalam penelitian ini menggunakan pendekatan komprehensif dengan metrik akurasi, *presicion*, *recall* dan *f1-score* baik saat proses pelatihan dan validasi ataupun tahap akhir dengan data pengujian. Pada tabel dibawah ini akan didapatkan data *classification report* berikut.

=== Evaluasi EfficientNetB0 ===					=== Evaluasi Vision Transformer ===				
Akurasi: 0.9877862595419847					Akurasi: 0.9984732824427481				
Precision: 0.9855492073367932					Precision: 0.9985611510791367				
Recall: 0.9873797893260845					Recall: 0.9965517241379309				
F1-Score: 0.9864477596758869					F1-Score: 0.9975388479045677				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.99	0.99	0.99	210	0	1.00	1.00	1.00	210
1	0.99	1.00	0.99	307	1	1.00	1.00	1.00	307
2	0.98	0.98	0.98	276	2	0.99	1.00	1.00	276
3	1.00	0.99	0.99	401	3	1.00	1.00	1.00	401
4	0.97	0.98	0.98	116	4	1.00	0.98	0.99	116
accuracy			0.99	1310	accuracy			1.00	1310
macro avg	0.99	0.99	0.99	1310	macro avg	1.00	1.00	1.00	1310
weighted avg	0.99	0.99	0.99	1310	weighted avg	1.00	1.00	1.00	1310

Gambar 4. Hasil Classification Report

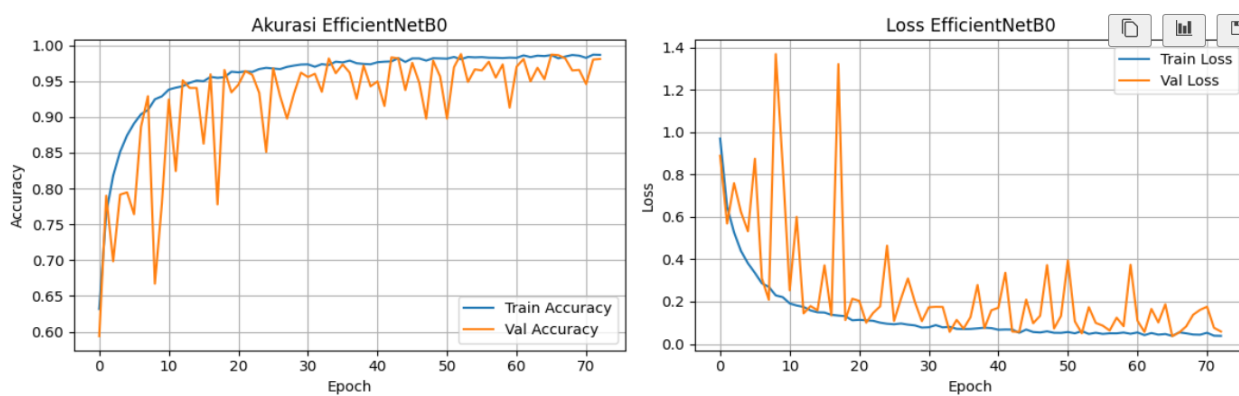
Dari Gambar 4 dapat diketahui performa klasifikasi terhadap lima kelas motif batik Kampung Giriloyo. Model EfficientNet menunjukkan akurasi yang tinggi sebesar 0.99, dengan *presicion*, *recall*, dan *f1-score*, *macro avg* dan *weighted avg* juga berada pada angka 0.99. Hal ini dapat diartikan bahwa EfficientNet mampu mengklasifikasikan secara akurat dan seimbang terhadap seluruh kelas meskipun terdapat penurunan performa pada kelas dengan jumlah data yang lebih sedikit. Kemudian pada model *Vision Transformer* mencatatkan performa yang lebih unggul dengan akurasi terbaik sebesar 1.00, serta nilai *presicion*, *recall*, dan *f1-score* yang mencapai 1.00 pada *macro avg* dan *weighted avg*. Seluruh kelas dalam model ini diklasifikasikan dengan sangat tepat meskipun *presicion* kelas ke-2 mendapatkan 0.99 dan *recall* kelas ke-4 hanya 0.98, namun hal ini tidak berdampak signifikan terhadap kinerja keseluruhan model. Maka dengan ini dapat dilihat hasil *Vision Transformer* dengan kemampuan generalisasi dan ekstraksi fitur visual motif batik dapat lebih kuat dibandingkan dengan EfficientNetB0 serta lebih teruji dalam menghadapi kontradiksi jumlah data antar kelas dan menjadikan model *Vision Transformer* terbaik pada studi ini.

3.6 Analisis Visualisasi dan Perbandingan hasil

Pada bagian ini akan menjelaskan hasil evaluasi yang akan ditampilkan dalam visualisasi grafik, *confusion matrix* dan tabel serta diagram perbandingan performa kedua model.

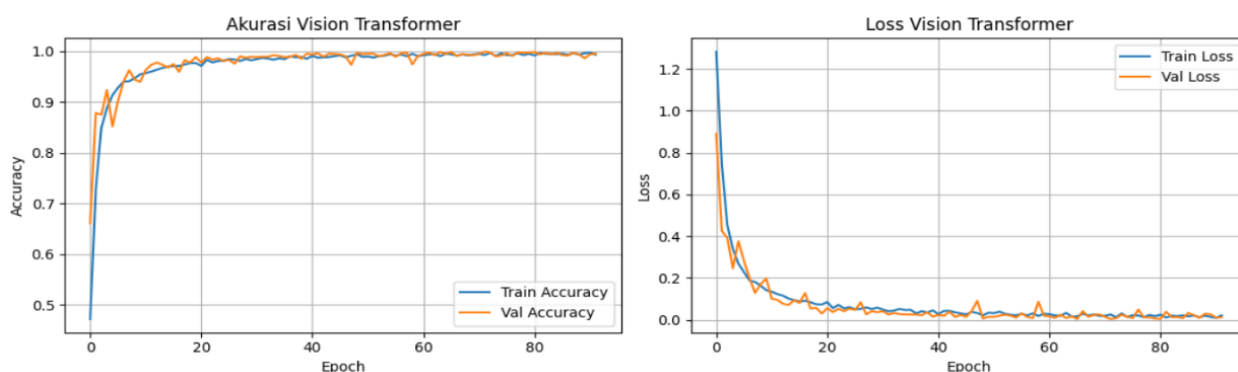
3.6.1 Grafik Akurasi dan Loss

Hasil evaluasi performa pelatihan dari model EfficientNetB0 dan *Vision Transformer* akan dijelaskan dan disajikan dalam grafik akurasi dan *loss* selama 100 epoch pelatihan. Grafik ini membantu menunjukkan efektivitas dan stabilitas model dalam proses pembelajaran pada data motif batik yang digunakan. Dibawah ini pada Gambar 5 dan Gambar 6 akan menampilkan grafik akurasi dan *loss* dari model EfficientNetB0 dan model *Vision Transformer*.



Gambar 5. Grafik Akurasi dan Loss Model EfficientNetB0

Analisis grafik akurasi dan *loss* dari model EfficientNetB0 dan *Vision Transformer* menunjukkan hasil yang sangat baik. Model menunjukkan peningkatan akurasi dan penurunan nilai *loss* seiring bertambahnya jumlah epoch. Namun terdapat perbedaan signifikan terhadap karakteristik pembelajaran kedua model tersebut yang mempengaruhi performa akhir dan efisiensi pelatihan. Pada grafik Gambar 5 dan Gambar 6, sumbu x menunjukkan jumlah epoch yang telah dijalankan dan sumbu y menunjukkan nilai akurasi dari tiap epoch. Berdasarkan gambar 5 pada model EfficientNetB0 menunjukkan tren peningkatan akurasi stabil pada data pelatihan, sementara akurasi validasi terlihat fluktuatif. Meski demikian, model tetap dapat mencapai akurasi validasi sebesar 96% pada sekitar epoch ke-30. Namun grafik *loss* validasi menunjukkan fluktuatif yang cukup tajam yaitu mencapai 1.4 sebelum stabil mendekati nilai yang lebih rendah. Nilai *loss* akhir dari EfficientNetB0 yaitu 0.08 pada pelatihan dan 0.12 pada validasi, hal ini mengindikasikan adanya sedikit kesenjangan dalam kemampuan model mempelajari data pelatihan serta generalisasi pada data validasi.



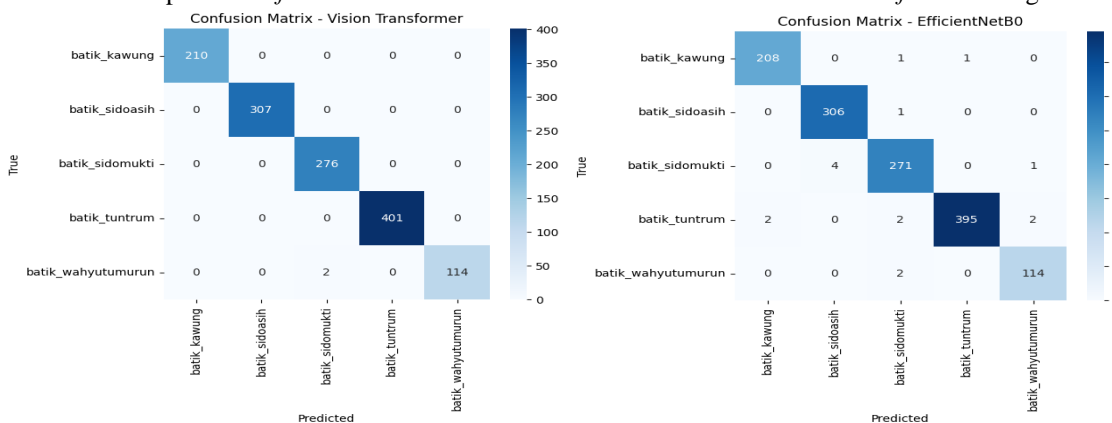
Gambar 6. Grafik Akurasi dan Loss Model Vision Transformer

Berbeda dengan Gambar 6 yang menunjukkan bahwa model *Vision Transformer* menghasilkan grafik akurasi yang jauh lebih stabil baik pada data pelatihan maupun validasi. Model ini mencapai titik pusat lebih cepat yaitu pada epoch ke-15 dengan nilai akurasi validasi sebesar 99%. Grafik *loss* juga menunjukkan adanya penurunan yang signifikan dan stabil dengan nilai *loss* akhir hanya 0.02 baik pada pelatihan maupun validasi. Hal ini menandakan kemampuan generalisasi model yang sangat baik serta minim terjadinya *overfitting*. Keunggulan *Vision Transformer* dalam hal stabilitas dan efisiensi pelatihan diperkuat oleh penelitian terdahulu berjudul “*Scaling Vision Transformer*” yang mengungkapkan adanya kecepatan konvergensi *Vision Transformer* dalam dataset berskala besar [19]. Selain itu penelitian “*Do Vision Transformers See Like Convolutional Neural Networks?*” yang juga menegaskan bahwa mekanisme *self-attention* pada arsitektur *transformer* memungkinkan model lebih efektif dibandingkan CNN konvensional sehingga mampu mempercepat proses pembelajaran fitur kompleks [20].

Keseluruhan analisis visualisasi ini menegaskan bahwa meskipun EfficientNetB0 masih mampu dalam mencapai akurasi validasi tinggi, performa *Vision Transformer* menunjukkan keunggulan yang lebih baik dalam akurasi, stabilitas pelatihan maupun efisiensi komputasi. Hal ini menjadikan *Vision Transformer* optimal dalam proses klasifikasi motif batik dengan pola kompleks.

3.6.2 Confusion Matrix

Bagian ini menampilkan hasil analisis akurasi tiap kelas berdasarkan *confusion matrix* dari dua model yang diuji. Gambar 7 akan menampilkan *confusion matrix* dari model EfficientNetB0 dan *Vision Transformer* sebagai berikut:



Gambar 7. Confusion Matrix Model EfficientNetB0 dan Vision Transformer

Berdasarkan hasil analisis pada Gambar 7, *Confusion Matrix* model EfficientNetB0 tepatnya gambar sebelah kiri terdapat 17 kesalahan klasifikasi yang tersebar di beberapa kelas. Sedangkan gambar sebelah kanan pada model *Vision Transformer* menunjukkan performa klasifikasi lebih baik dan minimnya kesalahan klasifikasi. Pada gambar tersebut model *Vision Transformer* hanya terdapat kesalahan klasifikasi pada kelas batik_wahyutumurun yang diklasifikasikan sebagai batik_sidomukti. Akurasi kedua model memiliki perbedaan signifikan dimana kelas batik_tuntrum memiliki jumlah sampel terbanyak 401/395. *Vision Transformer* mampu mengklasifikasikan 401 sampel dengan benar sementara EfficientNetB0 mengalami 6 kesalahan klasifikasi yaitu masing-masing 2 sampel yang diklasifikasikan sebagai batik_kawung, batik_sidomukti, dan batik_wahyutumurun. Kelas batik_wahyutumurun juga memiliki sampel paling sedikit yaitu 116 yang memperlihatkan performa identik antara kedua model dengan masing-masing mengalami 2 kesalahan klasifikasi terhadap kelas batik_sidomukti. Hal ini mengidentifikasi bahwa kedua model mengalami tantangan serupa dalam mengklasifikasikan kelas representasi sampel yang rendah. Hasil analisis pola klasifikasi EfficientNetB0 menunjukkan kesalahan tersebar dengan kecenderungan salah mengklasifikasikan batik_sidomukti sebagai batik_sidoasih sejumlah 4 sampel sedangkan *Vision Transformer* hanya

terjadi kesalahan pada pasangan batik_wahyutumurun dan batik_sidomukti. Hasil ini menegaskan bahwa model *Vision Transformer* lebih unggul dan konsisten dibandingkan dengan EfficientNetB0 dalam tugas klasifikasi citra motif batik secara mendalam. Klasifikasi yang lebih akurat serta konsisten dengan minimnya kesalahan juga dihasilkan oleh model tersebut. Pada EfficientNetB0 meskipun lebih rendah perbandingan hasil dengan *Vision Transformer* namun masih bisa menghasilkan performa yang cukup baik [21].

3.6.2 Perbandingan Performa Akurasi

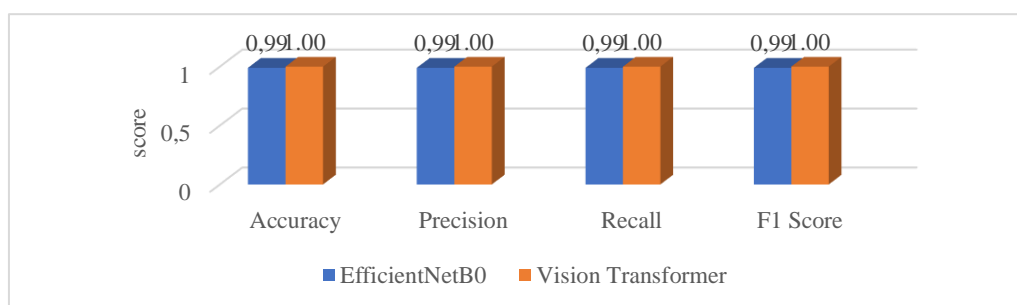
Pembahasan ini mendukung analisis performa kedua model dengan menyajikan perbandingan hasil evaluasi dalam bentuk tabel dan diagram batang yang menaggambarkan nilai akurasi, *presicion*, *recall* dan *f1-score* dari model EfficientNetB0 dan *Vision Transformer*. Hasil perbandingan performa akurasi dapat dilihat pada tabel 2 berikut.

Tabel 2. Perbandingan Performa Model EfficientNetB0 dan *Vision Transformer*

Model	Accuracy	Precision	Recall	F1-score
EfficientNetB0	98.78%	98.55%	98.74%	98.64%
<i>Vision Transformer</i> (ViT)	99.85%	99.86%	99.66%	99.75%

Pada Tabel 2 menunjukkan EfficientNetB0 juga memiliki hasil yang sangat baik dan mampu untuk dijadikan perbandingan yaitu *accuracy* 98.78%, *precision* sebesar 98.55%, *recall* 98.74% dan *f1-score* sebesar 98.64%. Meskipun menghasilkan nilai yang sangat baik, model *Vision Transformer* memperoleh hasil yang lebih unggul dibandingkan EfficientNetB0 dengan hasil akurasi mencapai 99.85%, *precision* senilai 99.86%, *recall* mendapatkan hasil 99.66% dan *f1-score* dengan hasil 99.75%. Hasil ini menyimpulkan bahwa tingkat akurasi sangat tinggi dan hampir sempurna baik dalam konteks klasifikasi general maupun per kelas. Berdasarkan laporan klasifikasi perolehan nilai *precision* dan *recall* diatas 99% menunjukkan bahwa model *Vision Transformer* mampu mengenali seluruh kelas motif dengan sangat akurat dan memiliki keseimbangan performa antar kelas yang sangat baik dibandingkan model EfficientNetB0 walaupun hasilnya juga <95%. Performa akurasi yang sangat tinggi menafsirkan bahwa keunggulan arsitektur *transformer* dalam memahami informasi visual dari citra serta mampu mengestrak representasi dari seluruh gambar pendekatan *self-attention* sehingga dapat menangkap pola dari motif batik yang cenderung kompleks, simetris dan terdapat pola berulang. Selain dari segi akurasi, efisiensi waktu inferensi juga diuji dalam penelitian ini sebagai bagian dari analisis model. Pengujian dengan 32 gambar uji menghasilkan waktu inferensi model EfficientNet sebesar 0.1924 detik sedangkan *Vision Transformer* membutuhkan waktu 0.1896 detik. Hal ini menunjukkan *Vision Transformer* tidak hanya unggul dalam akurasi tetapi juga tetap efisien waktu pada skenario inferensi batch kecil dan efisiensi ini menjadi nilai tambah yang signifikan.

Berikut informasi hasil perbandingan performa model yang terdiri dari *Accuracy*, *Precision*, *Recall*, dan *F1 Score* juga disajikan dalam *bar chart* atau diagram batang seperti Gambar 8 dibawah ini:



Gambar 8. Bar Chart Perbandingan Performa Model EfficientNetB0 dan *Vision Transformer*

Berdasarkan Gambar 8, model EfficientNetB0 menunjukkan performa yang sangat baik dengan nilai *accuracy*, *precision*, *recall* dan *f1 score* masing-masing nilai yaitu 0.99. Meski hasil menunjukkan tingkat akurasi yang tinggi dan stabil, model *Vision Transformer* memperoleh performa yang lebih unggul dengan skor sempurna pada keempat metrik evaluasi yaitu *accuracy* 1.00, *precision* sebesar 1.00, *recall* 1.00 dan *f1 score* 1.00. Nilai dengan selisih yang kecil namun konsisten ini menunjukkan bahwa *Vision Transformer* memiliki keunggulan dalam hal akurasi dan keseimbangan klasifikasi antar kelas. Kemampuan arsitektur *transformers* dalam menangkap pola visual melalui mekanisme *self-attention* memungkinkan model ini mengidentifikasi fitur penting secara menyeluruh terutama dalam konteks citra motif batik yang memiliki pola kompleks, berulang dan simetris. Maka dari itu, meskipun kedua model menunjukkan nilai yang sangat tinggi namun terdapat model yang lebih unggul baik dari segi akurasi maupun konsistensi hasil klasifikasi yaitu *Vision Transformer* dan kemudian menjadikannya sebagai pilihan yang optimal dalam studi ini.

4. KESIMPULAN

Kesimpulan dari hasil penelitian ini menunjukkan bahwa model *Vision Transformer* (ViT) memiliki keunggulan dibandingkan EfficientNetB0 dalam analisis perbandingan akurasi pada klasifikasi motif batik di Kampung Batik

Giriloyo. Model *Vision Transformer* berhasil mencapai hasil akurasi validasi sebesar 99.69% dan akurasi pengujian akhir sebesar 99.85%. Pada model EfficientNetB0 juga menunjukkan performa yang baik juga dengan akurasi validasi tertinggi 98.09% dan akurasi pengujian akhir sebesar 98.78%. Hasil ini diikuti dengan waktu inferensi yang berbeda pula yaitu pengujian dengan 32 gambar uji menghasilkan waktu inferensi model EfficientNetB0 sebesar 0.1924 detik sedangkan *Vision Transformer* membutuhkan waktu 0.1896 detik. Hasil evaluasi dan visualisasi metrix juga mengindikasikan bahwa *Vision Transformer* lebih unggul dalam klasifikasi motif batik yang kompleks dan memiliki kemiripan visual yang tinggi. Meskipun demikian, keterbatasan penelitian ini terdapat pada sumber daya komputasi tanpa dukungan GPU serta keterbatasan variasi jumlah data untuk beberapa kelas yang dapat memengaruhi stabilitas model. Selain itu, perlunya perbandingan analisis akurasi dengan model lain juga diperlukan untuk mendapatkan perbandingan yang lebih variatif. Maka dari itu, dapat disimpulkan bahwa *Vision Transformer* mampu menjadi model yang unggul dari model EfficientNetB0 pada CNN konvensional dalam klasifikasi motif batik khususnya di Kampung Batik Giriloyo baik segi akurasi maupun efisiensi. Temuan ini diharapkan dapat terus dikembangkan melalui sistem digitalisasi batik berbasis kecerdasan buatan yang akurat, efisien, adaptif untuk pelestarian budaya di era digital saat ini.

REFERENCES

- [1] R. Andrian, R. Taufik, D. Kurniawan, A. S. Nahri, and H. C. Herwanto, "Lampung Batik Classification Using AlexNet, EfficientNet, LeNet and MobileNet Architecture," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 11, pp. 930–935, 2024, doi: 10.14569/IJACSA.2024.0151191.
- [2] S. Aras, A. Setyanto, and Rismayani, "Deep Learning Untuk Klasifikasi Motif Batik Papua Menggunakan EfficientNet dan Transfer Learning," *Insect (Informatics Secur. J. Tek. Inform.)*, vol. 8, no. 1, pp. 11–20, 2022, doi: 10.33506/insect.v8i1.1865.
- [3] R. R. Karim, "Implementasi Klasifikasi Senjata Tradisional Jawa Barat Menggunakan Convolutional Neural Network (CNN) Dengan Metode Transfer Learning," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4166.
- [4] H. He, C. Wilson, T. T. Nguyen, and J. Dalins, "Sensitive Image Classification by Vision Transformers," *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, pp. 1704–1711, 2024, doi: 10.1109/SMC54092.2024.10831156.
- [5] M. Hayat, N. Ahmad, A. Nasir, and Z. A. Tariq, "Hybrid Deep Learning EfficientNetV2 and Vision Transformer (EffNetV2-ViT) Model for Breast Cancer Histopathological Image Classification," *IEEE Access*, vol. 12, no. December, pp. 184119–184131, 2024, doi: 10.1109/ACCESS.2024.3503413.
- [6] C. Testagrose, M. Shabbir, B. Weaver, and X. Liu, "Comparative Study Between Vision Transformer and EfficientNet on Marsh Grass Classification," *Proc. Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS*, vol. 36, 2023, doi: 10.32473/flairs.36.133132.
- [7] R. Pucci, V. J. Kalkman, and D. Stowell, "Comparison between transformers and convolutional models for fine-grained classification of insects," Arxiv, pp. 1–7, 2023, [Online]. Available: <http://arxiv.org/abs/2307.11112>
- [8] M. T. D. Putra *et al.*, "Batiknet: Batik Classification-based Management Application for Inexperienced User," *Int. J. Informatics Vis.*, vol. 8, no. 4, pp. 2411–2418, 2024, doi: 10.62527/joiv.8.4.3086.
- [9] M. Sulhan, "Perbandingan Metode Naïve Bayes Dengan SVM Pada Analisis Sentimen Aplikasi Pemesanan Tiket Kapal Ferizy," *BITS*, vol. 6, no. 4, pp. 0–9, 2025, doi: 10.47065/bits.v6i4.6715.
- [10] T. A. Bowo, H. Syaputra, and M. Akbar, "Penerapan Algoritma Convolutional Neural Network Untuk Klasifikasi Motif Citra Batik Solo," *J. Softw. Eng. Ampera*, vol. 1, no. 2, pp. 82–96, 2020, doi: 10.51519/journalsea.v1i2.47.
- [11] B.-K. Ruan, H.-H. Shuai, and W.-H. Cheng, "Vision Transformers: State of the Art and Research Challenges," Arxiv, no. 2, 2022, [Online]. Available: <http://arxiv.org/abs/2207.03041>
- [12] P. Khairunnisa, W. E. Putra, and W. Yitong, "Convolutional Neural Networks Using EfficientNetB0 Architecture and Hyperparameters on Skin Disease Classification," *Predatecs*, vol. 2, no. 2, pp. 127–137, 2025, doi: 10.57152/predatecs.v2i2.1569.
- [13] J. Mauricio, I. Domingues, and J. Bernardino, "Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review," *Appl. Sci.*, vol. 13, no. 9, 2023, doi: 10.3390/app13095521.
- [14] M. A. Salim, "Perbandingan Convolutional Neural Network dan SVM untuk Klasifikasi Penyakit Daun Padi," *J. Teknol. Dan Sist. Inf.*, vol. 3, no. 1, pp. 117–125, 2021.
- [15] M. Telçeken, D. Akgun, and S. Kacar, "An Evaluation of Image Slicing and YOLO Architectures for Object Detection in UAV Images," *Appl. Sci.*, vol. 14, no. 23, pp. 1–16, 2024, doi: 10.3390/app142311293.
- [16] J. Sadaiyandi, P. Arumugam, A. K. Sangaiah, and C. Zhang, "Stratified Sampling-Based Deep Learning Approach to Increase Prediction Accuracy of Unbalanced Dataset," *Electron.*, vol. 12, no. 21, pp. 1–16, 2023, doi: 10.3390/electronics12214423.
- [17] A. Nazarkar, H. Kuchulakanti, C. S. Paidimari, and S. Kulkarni, *Impact of Various Data Splitting Ratios on the Performance of Machine Learning Models in the Classification of Lung Cancer*, vol. 1. Atlantis Press International BV, 2023. doi: 10.2991/978-94-6463-252-1_12.
- [18] E. Setia Budi, A. Nofriyaldi Chan, P. Priscillia Alda, and M. Arif Fauzi Idris, "RESOLUSI: Rekayasa Teknik Informatika dan Informasi Optimasi Model Machine Learning untuk Klasifikasi dan Prediksi Citra Menggunakan Algoritma Convolutional Neural Network," *Media Online*, vol. 4, no. 5, p. 509, 2024, [Online]. Available: <https://djournals.com/resolusi>
- [19] X. Zhai, A. Kolesnikov, N. Houlsby, and L. Beyer, "Scaling Vision Transformers," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 12094–12103, 2022, doi: 10.1109/CVPR52688.2022.01179.
- [20] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do Vision Transformers See Like Convolutional Neural Networks?," *Adv. Neural Inf. Process. Syst.*, vol. 15, no. NeurIPS, pp. 12116–12128, 2021.
- [21] F. Y. Tember, I. Najiyah, "Klasifikasi Motif Batik Jawa Barat menggunakan Convolutional Neural Network Classification of West Java Batik Motifs Using Convolutional Neural Network," *Jurnal Sistemasi*, vol. 12, pp. 282–292, 2023, doi: 10.32520/stmsi.v12i2.2259