

ANALISIS PENGARUH AUGMENTASI DATA PADA ARSITEKTUR CNN DAN VISION TRANSFORMER UNTUK KLASIFIKASI MOTIF BATIK

Refando Nofiantoro, Tri Aristi Saputri

Teknik Informatika, Fakultas Teknologi Bisnis dan Sains, Universitas Dharma Wacana
Jl. Kenanga No.3, Mulyojati, Kec. Metro Barat, Kota Metro, Lampung
refandonofiantoro64@gmail.com

ABSTRAK

Klasifikasi citra Batik Nitik termasuk kategori *Fine Grained Visual Classification* (FGVC) karena tingginya kemiripan visual pada 60 kelas motifnya. Tantangan utama dalam klasifikasi ini adalah keterbatasan dataset orisinal yang hanya berjumlah 960 citra, sehingga sangat menyulitkan arsitektur *Vision Transformer* (ViT) yang memiliki karakteristik *data hungry*. Penelitian ini bertujuan menganalisis pengaruh augmentasi data terhadap kinerja *Convolutional Neural Network* (CNN) dan ViT dalam mengenali motif tersebut. Metode yang diusulkan adalah strategi Augmentasi Dua Tahap (ekspansi luring menjadi 9.600 citra dipadukan manipulasi spasial *on the fly*) dengan pengujian melalui Skenario 1 (Augmentasi Standar) dan Skenario 2 (Augmentasi Dua Tahap). Hasil pengujian Skenario 1 menunjukkan CNN gagal belajar (akurasi 1,67%) dan ViT mengalami *underfitting* (31,11%). Penerapan Skenario 2 berhasil meningkatkan akurasi CNN menjadi 63,44% yang tertahan oleh batas arsitekturalnya, sementara ViT mendominasi secara mutlak di angka 97,40%. Disimpulkan bahwa mekanisme *self attention* ViT lebih tangguh dari filter konvolusi CNN untuk FGVC, dengan prasyarat mutlak ketersediaan augmentasi data berskala masif.

Kata kunci : *Augmentasi Data, Batik Nitik, Convolutional Neural Network, Fine Grained Visual Classification, Vision Transformer*

1. PENDAHULUAN

Batik merupakan warisan budaya Indonesia yang telah mendapatkan pengakuan dunia oleh UNESCO sebagai *Masterpiece of Oral and Intangible Heritage of Humanity*. Salah satu varian batik yang memiliki nilai filosofis dan visual yang unik adalah Batik Nitik. Motif Batik Nitik merepresentasikan keragaman budaya Indonesia dengan tingkat kerumitan pola yang sangat tinggi dibandingkan motif makro seperti parang atau flora. Karakteristik utamanya yang tersusun dari akumulasi titik-titik geometris repetitif menempatkan klasifikasi Batik Nitik ke dalam kategori *Fine Grained Image Classification* (FGIC). Pada ranah ini, model dituntut harus mampu membedakan variasi antar kelas (*inter class*) yang sangat tipis sekaligus tetap andal terhadap fluktuasi dalam satu kelas (*intra class*) akibat perbedaan sudut pengambilan gambar maupun tekstur kain [1].

Dalam ranah Teknik Informatika, otomatisasi identifikasi terhadap 60 kelas Batik Nitik menggunakan *Deep Learning* menjadi instrumen vital bagi pelestarian warisan budaya, di mana arsitektur *Convolutional Neural Network* (CNN) dan inovasi *Vision Transformer* (ViT) menjadi metode mutakhir yang diandalkan untuk tugas tersebut.

Kompleksitas dan keragaman pola *Fine Grained* ini pada dasarnya telah menjadi hambatan signifikan bagi model konvolusional tradisional (CNN) untuk melakukan identifikasi secara akurat [2].

Meskipun saat ini CNN tetap menjadi standar utama karena kemampuannya dalam mengekstraksi fitur tekstur secara mendetail melalui mekanisme *inductive bias* [3], muncul arsitektur ViT yang mengadopsi mekanisme *self attention* untuk

memodelkan ketergantungan spasial secara global, sebuah kapabilitas yang sering kali menjadi kelemahan CNN [4], [5].

Namun, efektivitas ViT sangat terhambat oleh fenomena "*Data Hunger*", di mana performanya cenderung menurun drastis pada dataset kecil. Hal ini menjadi permasalahan krusial karena dataset Batik Nitik 960 secara publik hanya menyediakan 16 sampel asli per kelasnya. Kelangkaan data (*data scarcity*) yang ekstrem ini menjadi penghalang utama bagi model dalam mencapai stabilitas akurasi.

Berpijak pada permasalahan tersebut, penelitian komparatif ini bertujuan untuk mengevaluasi dan membandingkan secara empiris ketangguhan arsitektur CNN dan ViT dalam mengenali 60 variasi kelas motif Batik Nitik yang sangat identik. Secara spesifik, eksperimen ini ditujukan untuk membuktikan apakah manipulasi data sintetis mampu mengompensasi ketiadaan bias spasial pada ViT sekaligus mengatasi kelemahan *data hungry* nya. Selain itu, tujuan lainnya adalah untuk membandingkan seberapa efisien kemampuan ekstraksi fitur lokal pada *filter* CNN dalam menangani karakteristik unik dan kerumitan pola Batik Nitik saat dihadapkan pada variasi data berskala besar.

Sebagai rencana penyelesaian masalah atas kelangkaan data tersebut, penelitian ini mengusulkan strategi perluasan dataset dan eksplorasi teknik augmentasi (*Data Augmentation*) untuk meningkatkan keandalan model [6].

Rencana ini akan diimplementasikan melalui dua skenario eksperimen utama, yakni pengujian pada kondisi menggunakan dataset asli yang terbatas dan pengujian komparatif menggunakan dataset yang telah

diperluas secara masif. Melalui tahapan penyelesaian ini, diharapkan dapat ditemukan solusi arsitektur klasifikasi yang paling optimal dan *robust* dalam memetakan kerumitan fitur tekstur tradisional.

2. TINJAUAN PUSTAKA

2.1. Convolutional Neural Network (CNN)

CNN merupakan arsitektur *deep learning* yang dirancang khusus untuk memproses data berbentuk *grid* seperti citra digital. CNN bekerja dengan mengekstraksi fitur secara hierarkis, mulai dari fitur sederhana seperti tepi dan garis hingga fitur kompleks seperti tekstur [6].

Keunggulan utama CNN terletak pada *inductive bias* berupa lokalitas dan *translation invariance*, yang memungkinkannya mengenali pola tekstur repetitif secara efisien meski dengan jumlah data yang relatif terbatas [1].

Secara struktural, proses ekstraksi fitur ini didukung oleh penggunaan lapisan konvolusi (*convolutional layer*) dan lapisan *pooling* yang berfungsi mereduksi dimensi spasial tanpa membuang informasi visual yang penting. Pada kasus klasifikasi citra bertekstur rumit seperti motif batik, ketepatan penyusunan arsitektur CNN ini menjadi sangat krusial untuk memastikan model mampu membedakan variasi pola visual yang sangat kompleks [2].

Selain itu, keandalan CNN dalam menangkap detail lokal menjadikannya sangat ideal untuk tugas *fine grained visual classification* (FGVC), sehingga efektivitas ekstraksi fiturnya mampu mencapai akurasi yang maksimal dalam mengenali tekstur batik yang saling berdekatan [3].

2.2. Vision Transformer (ViT)

Berbeda dengan CNN, ViT mengadaptasi mekanisme *self attention* dari pemrosesan bahasa alami untuk visi komputer. Citra dibagi menjadi beberapa bagian kecil (*patches*) yang diproses secara *linear*. Mekanisme *multi head attention* pada ViT memungkinkan model untuk menangkap hubungan spasial global secara simultan, yang seringkali menjadi kelemahan pada CNN konvensional [5].

Namun, ViT cenderung memerlukan dataset yang jauh lebih besar (*data hungry*) karena ketiadaan bias spasial bawaan seperti pada konvolusi [4].

Untuk mengatasi masalah kelangkaan data tersebut, pendekatan *transfer learning* sangat disarankan, di mana ViT memanfaatkan bobot fitur pra latih (*pre trained weights*) dari dataset berskala masif untuk memandu model beradaptasi pada tugas klasifikasi yang spesifik [7].

Lebih lanjut, integrasi arsitektur ini dengan teknik augmentasi data terbukti krusial untuk meningkatkan *robustness* serta kemampuan generalisasi model, sehingga meminimalisasi risiko *overfitting* akibat data yang terbatas [8].

Melalui kombinasi strategi *fine tuning* dan manipulasi data dinamis ini, mekanisme *self attention* pada ViT dapat dieksploitasi secara optimal untuk

memodelkan struktur repetitif dan detail granular pada motif batik secara menyeluruh.

2.3. Augmentasi Data

Augmentasi Data: Teknik ini merupakan strategi untuk memperluas variasi dan *volume* dataset secara artifisial. Augmentasi sangat krusial dalam klasifikasi kain tradisional untuk meningkatkan ketangguhan model terhadap variasi sudut pengambilan gambar, pencahayaan, dan tekstur [3].

Melalui manipulasi seperti rotasi, *flipping*, dan *zooming*, model dapat belajar lebih baik tanpa mengalami *overfitting* pada dataset yang kecil. Dalam penerapannya, manipulasi data secara dinamis (*on the fly*) selama proses pelatihan memberikan keuntungan komputasional karena model terus dilatih dengan variasi citra baru pada setiap *epoch* tanpa perlu memproduksi dan menyimpan data buatan secara fisik. Pendekatan intervensi visual ini merupakan solusi yang sangat efektif dalam mengatasi kendala kelangkaan data (*data scarcity*) pada objek-objek spesifik yang memiliki jumlah sampel awal sangat terbatas [9].

Di samping itu, perlakuan augmentasi ini tidak hanya berfungsi sebagai pencegah *overfitting*, tetapi juga menjadi instrumen pengujian yang krusial untuk mengevaluasi dan membandingkan secara empiris bagaimana arsitektur konvensional (CNN) dan arsitektur berbasis transformer (ViT) bereaksi terhadap distorsi spasial yang sama [10].

2.4. Karakteristik Visual dan Klasifikasi Batik Nitik

Karakteristik Visual dan Klasifikasi Batik Nitik: Batik Nitik merupakan salah satu varian motif batik tradisional yang secara etimologis berasal dari kata "titik". Berbeda dengan motif batik klasik pada umumnya yang didominasi oleh garis lengkung dan ornamen dinamis, pola dasar Batik Nitik secara eksklusif tersusun dari akumulasi ribuan titik geometris yang berulang secara presisi untuk meniru tekstur jalinan benang pada kain tenun. Dalam ranah *Computer Vision*, klasifikasi citra Batik Nitik menghadirkan tantangan komputasi yang berat dan secara teknis dikategorikan ke dalam permasalahan *Fine Grained Visual Classification* (FGVC) akibat fluktuasi intra-kelas dan tingginya kemiripan antar-kelas [1].

Saat mengklasifikasikan 60 variasi kelas motif Nitik yang berbeda, perbedaan visual antar motif sering kali hanya terletak pada aspek mikroskopis seperti kerapatan titik, orientasi sudut, atau formasi makro geometrisnya. Karakteristik granular dan repetitif yang sangat kompleks inilah yang menuntut ketepatan dalam pemilihan arsitektur *deep learning* agar model memiliki kapabilitas ekstraksi fitur tekstur secara sangat mendalam dan presisi tinggi [2].

2.5. Metrik Evaluasi dan Analisis Kinerja Model

Metrik Evaluasi dan Analisis Kinerja Model: Mengingat fokus utama penelitian ini bertumpu pada analisis komparatif, diperlukan instrumen pengukuran matematis yang objektif untuk memvalidasi dan membandingkan kinerja arsitektur model terhadap intervensi augmentasi data secara mendalam [10].

Analisis kinerja model dalam penelitian ini dievaluasi melalui tiga indikator utama yang saling melengkapi. Pertama, Akurasi (*Accuracy*), yang berfungsi sebagai parameter fundamental untuk mengukur rasio persentase citra motif batik yang berhasil diprediksi secara benar oleh model dari keseluruhan populasi data uji. Kedua, Fungsi Kerugian (*Loss Function*) menggunakan *Categorical Crossentropy* untuk menghitung deviasi antara distribusi probabilitas hasil prediksi dengan label kebenaran aktual, di mana pemantauan *curva loss* sangat krusial untuk mengidentifikasi gejala kegagalan belajar model seperti *overfitting* atau *underfitting*. Ketiga, *Confusion Matrix*, yang merupakan matriks kontingensi untuk menyajikan rincian tabulasi silang antara prediksi model dan label aktual. Instrumen *Confusion Matrix* ini memungkinkan peneliti untuk melakukan analisis spesifik dalam mengidentifikasi kelas motif batik mana yang paling sering mengalami tumpang tindih prediksi (*misclassification*) oleh arsitektur *CNN* maupun *ViT* akibat kedekatan karakteristik visualnya.

2.6. Penelitian Terdahulu yang Terkait

Beberapa penelitian terdahulu telah membuktikan ketangguhan arsitektur *Convolutional Neural Network (CNN)* dalam menangani ekstraksi fitur citra bertekstur, khususnya pada pengenalan motif batik dan kain tradisional. [1] serta [2] secara meyakinkan menunjukkan bahwa arsitektur berbasis *CNN* seperti *ResNet* dan *MobileNet* sangat andal dalam mengklasifikasikan motif batik Indonesia yang memiliki kerumitan tinggi.

Temuan ini sejalan dengan penelitian [11] pada klasifikasi batik menggunakan *ResNet-50*, serta [6] yang sukses mengidentifikasi kelas motif kain Jumputan Palembang dengan metode *CNN* serupa. Untuk mencapai akurasi klasifikasi yang maksimal, arsitektur konvensional ini juga kerap dimodifikasi, seperti penggabungan ekstraksi fitur tekstur *GLCM* dan *CNN* oleh [3], hingga implementasi *MobileNet* yang diintegrasikan dengan sistem *chatbot Retrieval Augmented Generation (RAG)* oleh [12].

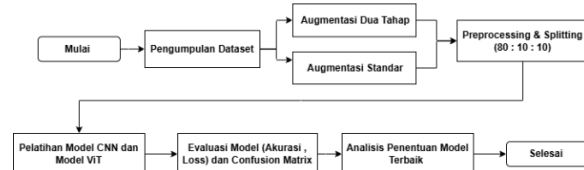
Seiring perkembangan tren kecerdasan buatan, arsitektur *Vision Transformer (ViT)* mulai dieksplorasi secara masif pada berbagai domain visi komputer. Keunggulan *ViT* dalam menangkap dependensi spasial jarak jauh telah dibuktikan keberhasilannya oleh [5] pada deteksi penyakit daun cabai, [4] pada klasifikasi sel darah putih, [13] untuk deteksi kematangan pisang, hingga [14] untuk pengenalan ekspresi wajah peserta didik. Meskipun memiliki keunggulan pemahaman fitur secara global, perdebatan terkait perbandingan

performa antara *ViT* dan *CNN* masih berlanjut. Studi komparasi yang dilakukan oleh [7] pada dataset citra medis, serta [15] pada klasifikasi kanker payudara, menyimpulkan bahwa arsitektur *CNN* kerap kali masih mampu mengungguli *ViT* pada kondisi dataset tertentu. Hal ini mengindikasikan adanya kebutuhan intervensi perlakuan data untuk dapat mengeksplorasi potensi *ViT* secara optimal. Teknik augmentasi data telah diakui secara luas sebagai solusi utama untuk mengatasi keterbatasan sampel sekaligus mencegah terjadinya *overfitting*.

Efektivitas augmentasi ini telah terbukti di berbagai disiplin ilmu, mulai dari pengenalan citra mobil oleh [16], identifikasi daun padi oleh [17]) dan daun tebu oleh [18], penerjemahan bahasa isyarat BISINDO oleh [19], pembacaan aksara Kawi pada prasasti logam oleh [9], hingga ranah prediksi teks kepribadian *MBTI* oleh [20].

Secara spesifik pada identifikasi motif batik, penelitian [21] serta membuktikan bahwa pemanfaatan augmentasi data secara signifikan sukses menaikkan akurasi klasifikasi model. Bagi arsitektur *transformer*, manipulasi data artifisial terbukti mampu mendongkrak ketahanan (*robustness*) model, sebagaimana ditunjukkan oleh [8] pada pengenalan visual masakan Padang. Analisis komparatif mendalam terkait reaksi arsitektur *CNN* dan *ViT* terhadap berbagai variasi augmentasi seperti yang dipelopori oleh [10] pada karakter Mandarin, menjadi landasan yang sangat kuat bagi riset ini. Berpijak pada 21 literatur tersebut, penelitian ini difokuskan untuk mengisi celah kajian dengan mengevaluasi secara empiris dampak penerapan strategi augmentasi data dua tahap terhadap efisiensi *ViT* dan *CNN* dalam menangani *Fine Grained Visual Classification* pada 60 kelas motif Batik Nitik.

3. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Dari alur tahapan penelitian pada gambar 1 mengilustrasikan tahapan metodologi secara sistematis yang dilakukan dalam penelitian ini. Proses diawali dengan tahap pengumpulan dataset citra motif Batik Nitik. Dataset yang telah terkumpul kemudian dikondisikan ke dalam dua skenario eksperimen yang berjalan paralel, yakni penerapan augmentasi standar dan augmentasi dua tahap guna menguji keandalan model. Selanjutnya, seluruh data dari kedua skenario tersebut memasuki fase pra pemrosesan (*preprocessing*) dan pembagian data (*data splitting*), di mana himpunan data dipartisi menjadi tiga subset utama dengan rasio 80% untuk data latih (*training*), 10% untuk data validasi (*validation*), dan 10% untuk

data uji (*testing*). Dataset yang telah dipersiapkan ini kemudian diumpankan ke dalam tahap pelatihan model secara komparatif antara arsitektur *Convolutional Neural Network (CNN)* dan *Vision Transformer (ViT)*. Setelah fase pelatihan rampung, kinerja dari masing-masing arsitektur dievaluasi secara komprehensif bersandar pada metrik akurasi, perhitungan nilai kerugian (*loss*), serta visualisasi *confusion matrix*. Seluruh rangkaian penelitian ini diakhiri dengan tahap analisis komparatif mendalam berdasarkan hasil evaluasi tersebut guna menentukan dan menyimpulkan model terbaik yang paling efisien dalam mengklasifikasikan citra *fine grained*.

3.1. Teknik Pengumpulan Data

Dataset citra yang digunakan dalam penelitian ini merupakan dataset sekunder yang diperoleh secara publik dari repositori *Mendeley Data* dengan nama rilis "Batik Nitik 960". Mengingat dataset awal citra Batik Nitik yang terkumpul hanya berjumlah 960 citra (terbagi ke dalam 60 kelas, dengan masing-masing 16 citra per kelas), jumlah ini sangat rentan memicu fenomena *overfitting* dan belum memenuhi standar kebutuhan arsitektur *Vision Transformer (ViT)* yang cenderung *data hungry*. Adapun rincian dari 60 kelas motif Batik Nitik yang diekstraksi dan diklasifikasikan dalam penelitian ini dijabarkan secara lengkap pada Tabel 1.

Tabel 1. Rincian 60 Kelas Motif Batik Nitik

Kelas Motif		
1. Sekar Kemuning	21. Mawur	41. Sekar Andhong
2. Sekar Liring	22. Sekar Tanjung	42. Gedhangan
3. Sekar Duren	23. Sekar Keben	43. Sekar Pala
4. Sekar Gayam	24. Sekar Srengenge	44. Klampok Arum
5. Sekar Pacar	25. Sekar Soka	45. Sekar Jali
6. Arumdalu	26. Sekar Nangka	46. Sekar Lintang
7. Sekar Srigadhing	27. Kawung Nitik	47. Sekar Kenanga
8. Kemukus	28. Sekar Kenthang	48. Sekar Jeruk
9. Sekar Gudhe	29. Sekar Pudhak	49. Sekar Mindi
10. Sekar Ketongkeng	30. Sekar Dlima	50. Tanjung Gunung
11. Brendhi	31. Krawitan	51. Sekar Kenikir
12. Cakar Ayam	32. Cinde Wilis	52. Sekar Blimbing
13. Sekar Menur	33. Sekar Mlathi	53. Sekar Pijetan
14. Sekar Tebu	34. Kuncup Kanthil	54. Sari Mulat
15. Sekar Manggis	35. Sekar Dangan	55. Sekar Mrica
16. Sekar Randhu	36. Sekar Sawo	56. Sekar Kepel
17. Worawari Rumpuk	37. Manggar	57. Truntum Kurung

Kelas Motif		
18. Sekar Dhuku	38. Sekar Cengkeh	58. Jayakusuma
19. Sekar Jagung	39. Sritaman	59. Rengganis
20. Jayakirana	40. Sekar Mundhu	60. Sekar Gambir

Oleh karena itu, penelitian ini mengimplementasikan Strategi Augmentasi Dua Tahap (*Two Stage Augmentation*) untuk memastikan volume dan variasi data terpenuhi secara optimal. Tahap Pertama: Augmentasi *Offline* (Peningkatan *Volume Dataset*) bertujuan untuk mengekspansi skala dataset secara fisik sebelum proses pelatihan dimulai. Melalui proses augmentasi *offline* ini, jumlah data meningkat secara signifikan dari 16 citra menjadi 160 citra per kelas, sehingga menghasilkan total dataset fisik sebanyak 9.600 citra. Tahap Kedua: Augmentasi *On The Fly* (Variasi Dinamis Spasial), untuk mencegah model sekadar menghafal 9.600 citra fisik yang telah dibuat, penelitian ini menambahkan lapisan augmentasi kedua secara dinamis (*on the fly*) menggunakan fungsi *ImageDataGenerator* saat proses pelatihan berlangsung. Parameter intervensi visual yang diterapkan pada tahap ini meliputi: Rotasi, *Zooming*, *Flipping*, *Shifting*, dan *Shearing*.

3.2. Pre-processing dan Pembagian Dataset (Data Splitting)

Sebelum memasuki tahap pelatihan model, dataset berukuran 9.600 citra tersebut melalui tahap pra pemrosesan (*pre processing*). Seluruh citra diubah ukurannya (*resize*) secara seragam menjadi dimensi 224×224 piksel untuk menyesuaikan standar input masukan pada arsitektur *CNN* dan *ViT*. Selanjutnya, dilakukan normalisasi nilai piksel dengan teknik *rescaling* ($1./255$) agar rentang nilai intensitas warna berada di antara 0 dan 1, yang berfungsi untuk mempercepat komputasi konvergensi model. Dataset kemudian dibagi menjadi tiga subset menggunakan modul *splitfolders* dengan rasio 80:10:10, yaitu 80% untuk data latih (*training*), 10% untuk data validasi (*validation*), dan 10% untuk data uji (*testing*).

3.3. Arsitektur Model dan Skenario Pelatihan

Penelitian ini membandingkan kinerja dua arsitektur *deep learning* yang beroperasi di bawah batasan lingkungan komputasi yang seragam. Untuk mengukur efisiensi komparatif, proses pelatihan dibatasi secara ketat maksimal 10 *epoch* dengan ukuran batch (*batch size*) sebesar 16 untuk kedua model. Skenario pemodelan mencakup: Arsitektur *CNN*: Dibangun dari awal (*from scratch*) tanpa menggunakan bobot pra latih. Arsitektur ini terdiri dari beberapa blok lapisan konvolusi (*Conv2D*) yang dipadukan dengan *Batch Normalization* dan *MaxPooling2D* untuk ekstraksi fitur hierarkis, diakhiri dengan *Flatten* dan *Fully Connected Layer (Dense)* menggunakan fungsi aktivasi *Softmax* untuk klasifikasi 60 kelas. *Optimizer* yang digunakan adalah Adam. Arsitektur *Vision Transformer (ViT)*: Dibangun

menggunakan pendekatan *Transfer Learning*. Lapisan representasi dasar yang berisi bobot pra-latih dari dataset berskala besar dibekukan (*frozen*), sehingga proses pembaruan bobot (*fine tuning*) hanya difokuskan pada lapisan klasifikasi akhir (*head layer*) untuk mengklasifikasikan 60 kelas Batik Nitik. Pengujian ini membagi pemodelan menjadi dua skenario utama, yakni Skenario 1 yang menggunakan dataset orisinal ditambah augmentasi *on the fly*, dan Skenario 2 yang menggunakan dataset hasil ekspansi *offline* 9.600 citra ditambah augmentasi *on the fly*.

3.4. Metrik Evaluasi Kinerja

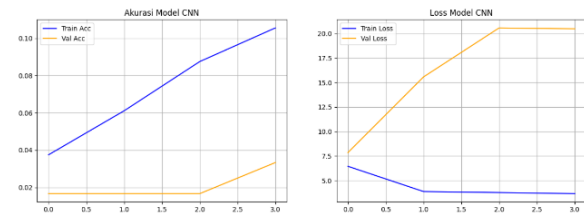
Kinerja klasifikasi dari kedua model dievaluasi secara kuantitatif menggunakan tiga instrumen pengukuran utama: Akurasi (*Accuracy*): Menghitung rasio persentase prediksi kelas motif yang benar dibandingkan dengan total keseluruhan data uji (*test set*). *Categorical Crossentropy Loss*: Berfungsi sebagai fungsi kerugian untuk mengevaluasi probabilitas kesalahan prediksi. Pemantauan kurva *loss* pada data latih dan validasi dilakukan untuk mendeteksi indikasi *underfitting* maupun *overfitting* selama durasi 10 *epoch*. *Confusion Matrix*: Evaluasi visual matriks kontingensi untuk memetakan secara detail distribusi tumpang tindih kesalahan prediksi (*misclassification*) antar kelas variasi Batik Nitik oleh masing-masing arsitektur. Perhitungan komprehensif pada matriks ini bertumpu pada empat elemen dasar, yakni *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Berdasarkan elemen tersebut, diekstraksi tiga metrik tambahan untuk melihat keseimbangan model: Presisi (*Precision*): Mengukur tingkat ketepatan prediksi kelas positif. *Recall* (Sensitivitas): Mengukur kemampuan model menemukan kembali seluruh citra kelas yang sebenarnya. *F1-Score*: Nilai rata-rata harmonik antara presisi dan recall yang krusial untuk evaluasi dataset multi kelas yang kompleks.

4. HASIL DAN PEMBAHASAN

4.1. Skenario Pengujian

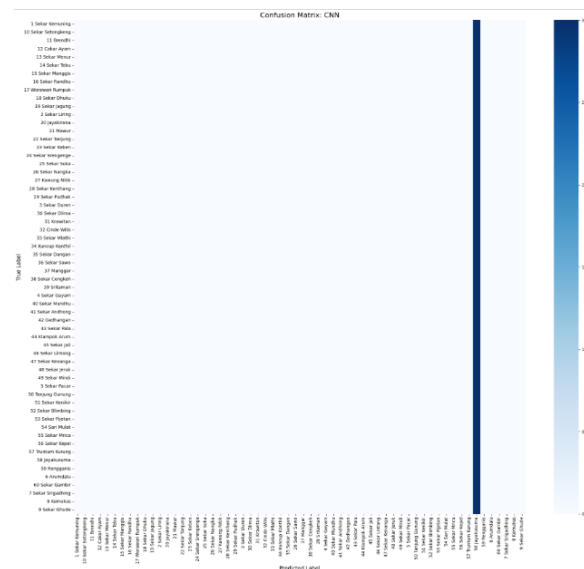
Penelitian ini berfokus pada analisis komparatif untuk mengukur sejauh mana pengaruh skala augmentasi data terhadap kinerja klasifikasi model *Convolutional Neural Network (CNN)* dan *Vision Transformer (ViT)*. Untuk mendapatkan kesimpulan yang objektif, pengujian dibagi menjadi dua skenario utama: Skenario 1 (Augmentasi Standar): Mengevaluasi model menggunakan dataset orisinal (960 citra) yang hanya diberikan augmentasi dinamis (*on the fly*) saat proses pelatihan berlangsung. Skenario 2 (Augmentasi Dua Tahap): Mengevaluasi model menggunakan dataset yang telah diekspansi secara luring (*offline*) menjadi 9.600 citra, yang kemudian dipadukan dengan augmentasi dinamis (*on the fly*). Kedua skenario tersebut dilatih secara seragam menggunakan rasio pembagian data 80:10:10 dan dibatasi maksimal 10 *epoch*.

4.2. Hasil Pengujian Skenario 1: Augmentasi Standar (Dataset 960 Citra + On The Fly) Kinerja Arsitektur CNN (Skenario 1)



Gambar 2. Kurva Loss dan Akurasi Model CNN Skenario 1

Berdasarkan hasil pengujian pada Skenario 1, model *CNN* menunjukkan kegagalan pembelajaran yang fatal. Akurasi pengujian (*test accuracy*) terpuruk di angka 1.67%. Secara matematis, angka 1.67% ini merepresentasikan probabilitas tebakan acak (*random guess*), di mana peluang menebak 1 kelas dari total 60 kelas motif adalah $1/60 = 0.0167$. Fenomena ini membuktikan bahwa meskipun *CNN* telah diberikan variasi augmentasi *on the fly* seperti rotasi dan *shearing*, arsitektur yang dibangun dari awal (*from scratch*) ini tetap gagal mengekstrak pola hierarkis akibat *volume* data dasar (960 citra) yang terlalu minim untuk mencakup 60 kelas yang kompleks.



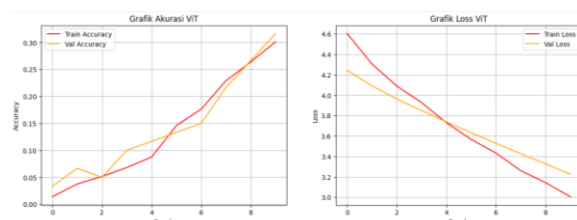
Gambar 3. *Confusion Matrix* Model CNN Skenario 1

Evaluasi matriks pada Gambar 4.2 mengonfirmasi hal tersebut. Model *CNN* gagal mendeteksi karakteristik antar kelas, menghasilkan *F1-Score* yang mendekati nol, akibat distribusi bobot yang tidak terkonvergensi.

Berdasarkan evaluasi terhadap laporan klasifikasi (*classification report*), kinerja model *CNN* pada skenario ini menunjukkan indikasi kegagalan pembelajaran yang fatal dengan tingkat akurasi keseluruhan hanya sebesar 2%. Analisis lebih mendalam pada distribusi metrik mengungkapkan terjadinya fenomena *mode collapse*, di mana model

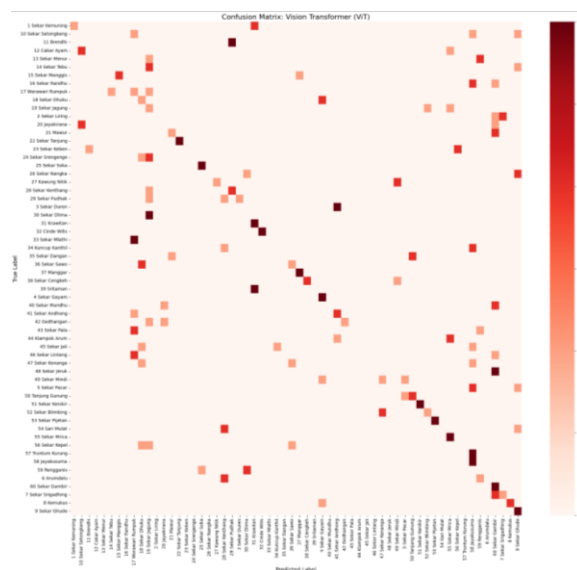
kehilangan kemampuan diskriminatifnya dan cenderung memprediksi secara buta seluruh citra uji ke dalam satu representasi kelas tunggal, yakni kelas "58 Jayakusuma". Fenomena ini secara empiris mengonfirmasi bahwa arsitektur *CNN* yang dilatih dari awal (*from scratch*) pada dataset *bervolume* minim gagal mengonversi bobot jaringannya.

4.3. Kinerja Arsitektur Vision Transformer (Skenario 1)



Gambar 4. Kurva Loss dan Akurasi Model ViT Skenario 1

Berbeda dengan *CNN*, arsitektur ViT pada Skenario 1 mampu bertahan dengan tingkat akurasi sebesar 31.11%. Pencapaian ini murni didorong oleh pendekatan *Transfer Learning*. Bobot pra latih membantu model mengenali bentuk dasar geometris, namun augmentasi *on the fly* pada 960 citra ternyata belum cukup untuk mengatasi sifat *data hungry* dari arsitektur *transformer* ini.

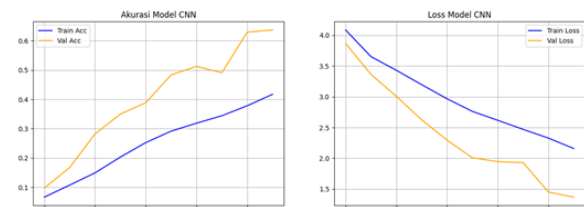


Gambar 5. Confusion Matrix Model ViT Skenario 1

Berdasarkan hasil evaluasi laporan klasifikasi, arsitektur *Vision Transformer* (ViT) pada skenario pembatasan data menunjukkan kinerja parsial dengan tingkat akurasi keseluruhan mencapai 31%. Meskipun demikian, distribusi metrik secara keseluruhan mengindikasikan bahwa model sedang mengalami fase *underfitting* yang parah. Fenomena ketidakseimbangan ini secara empiris memvalidasi karakteristik *data hungry* bawaan dari arsitektur *transformer*, di mana *volume* 960 citra orisinal yang

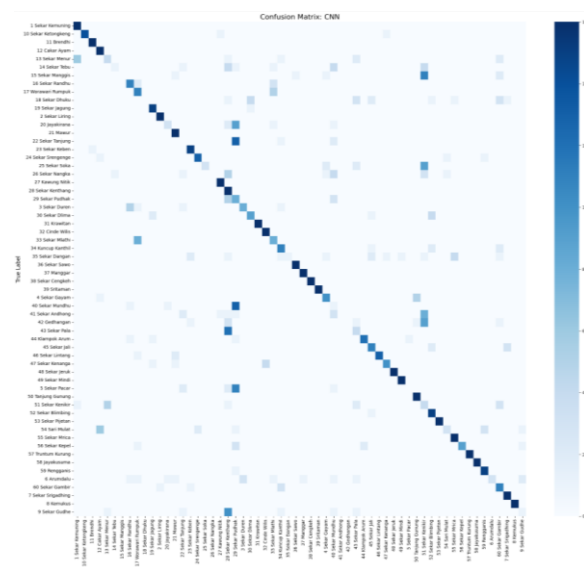
hanya dibantu augmentasi *on the fly* terbukti tidak memberikan cukup variasi informasi untuk mengkalibrasi mekanisme *Multi Head Self Attention* dalam membedakan 60 kelas motif *fine grained* yang sangat identik.

4.4. Hasil Pengujian Skenario 2: Augmentasi Dua Tahap (Dataset 9.600 Citra + On The Fly) Kinerja Arsitektur CNN (Skenario 2)



Gambar 6. Kurva Loss dan Akurasi Model CNN Skenario 2

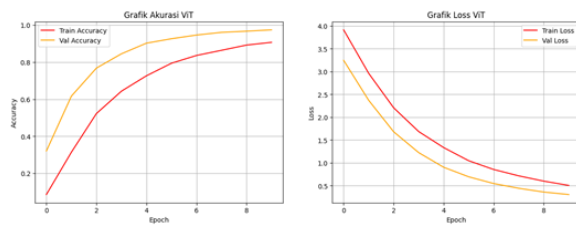
Penerapan augmentasi dua tahap terbukti menjadi penyelamat bagi model *CNN*. Pasokan 9.600 citra sebagai basis pengetahuan visual berhasil menyembuhkan kelumpuhan model, mendorong akurasi pengujian melonjak drastis dari 1.67% menjadi 63.44%.



Gambar 7. Confusion Matrix Model CNN Skenario 2

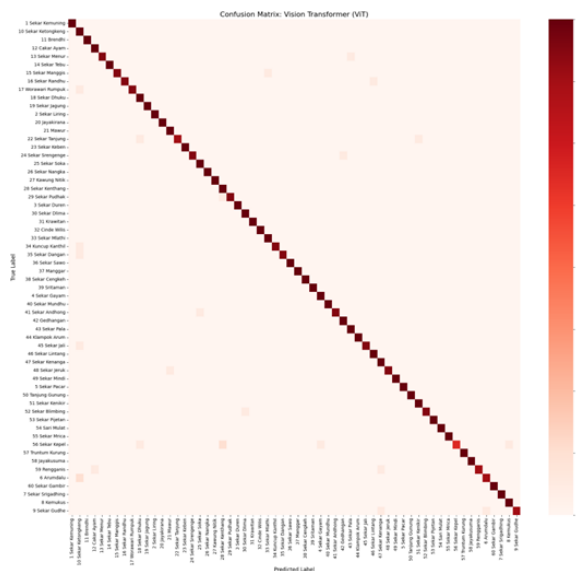
Berdasarkan hasil evaluasi laporan klasifikasi pada Skenario 2, penerapan strategi augmentasi data berskala masif terbukti memberikan dampak transformatif dengan mendongkrak akurasi model *CNN* secara signifikan menjadi 63%. Meskipun demikian, fluktuasi metrik antar kelas juga secara empiris mengonfirmasi bahwa *CNN* telah menyentuh batas efisiensi arsitekturalnya (*ceiling performance*). Fenomena ini menyimpulkan bahwa penambahan *volume* data semata tidak cukup untuk mengompensasi kelemahan intrinsik *CNN* dalam mempertahankan presisi konteks global saat mengklasifikasikan 60 variasi tekstur *Fine Grained* yang sangat identik.

4.5. Kinerja Arsitektur Vision Transformer (Skenario 2)



Gambar 8. Kurva Loss dan Akurasi Model ViT Skenario 2

Skenario 2 menjadi pembuktian superioritas absolut arsitektur ViT. Penggabungan 9.600 citra fisik dengan augmentasi *on the fly* secara sempurna memuaskan kebutuhan *data hungry* dari ViT. Akurasi ViT melonjak secara eksponensial mencapai 97.40%. Kurva *loss* menekuk tajam secara konvergen tanpa indikasi *overfitting*.



Gambar 9. Confusion Matrix Model ViT Skenario 2

Berdasarkan hasil evaluasi laporan klasifikasi pada Skenario 2, arsitektur Vision Transformer (ViT) menunjukkan dominasi performa yang mutlak dengan pencapaian akurasi keseluruhan sebesar 97%. Berbeda dengan model CNN yang tertahan oleh batas efisiensi arsitekturalnya, penerapan strategi augmentasi data berskala masif terbukti secara sempurna berhasil mengatasi karakteristik *data hungry* pada ViT. Hasil empiris ini secara konklusif menegaskan bahwa arsitektur berbasis *transformer* merupakan solusi komputasi yang jauh lebih tangguh untuk klasifikasi citra tekstur kompleks, dengan prasyarat mutlak berupa ketersediaan pasokan intervensi data spasial yang melimpah.

4.6. Analisis Pengaruh Augmentasi Data (Kesimpulan Komparasi)

Tabel 2. Rangkuman Komparasi Akurasi Kinerja Model

Arsitektur Model	Akurasi Skenario 1 (Augmentasi Standar)	Akurasi Skenario 2 (Augmentasi Dua Tahap)
Convolutional Neural Network (CNN)	1.67%	63.44%
Vision Transformer (ViT)	31.11%	97.40%

Data empiris dari pengujian ini menjawab tuntas rumusan masalah penelitian. Intervensi augmentasi data memiliki pengaruh yang sangat krusial, namun efektivitasnya sangat bergantung pada skala penerapannya. Augmentasi dinamis (*on the fly*) saja terbukti tidak cukup tangguh untuk menangani klasifikasi citra *Fine Grained* seperti Batik Nitik pada dataset yang sangat minim. Sebaliknya, strategi Augmentasi Dua Tahap yang diusulkan dalam penelitian ini memberikan dampak transformatif. Bagi CNN, strategi ini menyelamatkan model dari status gagal belajar (*random guessing*) menjadi model fungsional (63.44%). Sementara bagi Vision Transformer, strategi ini bertindak sebagai katalisator yang melepaskan potensi penuh mekanisme *self attention*, menyembuhkan masalah *underfitting*, dan mendorong akurasi klasifikasi menembus 97.40%.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian, evaluasi metrik, dan analisis komparatif yang telah diuraikan pada bab sebelumnya, penelitian ini menghasilkan kesimpulan untuk menjawab rumusan masalah sebagai berikut: Efektivitas Arsitektur CNN berbanding ViT pada Klasifikasi Batik Nitik: Arsitektur Vision Transformer (ViT) terbukti jauh lebih efektif dan superior dibandingkan Convolutional Neural Network (CNN) dalam mengenali 60 kelas motif Batik Nitik yang bersifat *Fine Grained*. Pada kondisi pasokan data yang masif, ViT mendominasi secara mutlak dengan pencapaian akurasi sebesar 97.40%, meninggalkan CNN yang tertahan pada batas efisiensinya di angka 63.44%. Keunggulan ini dipicu oleh mekanisme *Multi Head Self Attention* pada ViT yang mampu mempertahankan pemahaman konteks spasial secara global meskipun citra mengalami distorsi kemiringan dan rotasi, sedangkan *filter* konvolusi lokal pada CNN gagal mempertahankan presisinya saat dihadapkan pada variasi posisi motif yang rumit. Pengaruh Augmentasi Data terhadap Kinerja CNN dan ViT:

Penelitian ini membuktikan secara empiris bahwa intervensi augmentasi data skala masif (Strategi Dua Tahap) memberikan dampak transformatif yang sangat krusial bagi kedua arsitektur. Bagi CNN, strategi augmentasi ini bertindak sebagai penyelamat yang mengubah model dari status gagal belajar (*random guessing* di angka 1.67%) menjadi model

fungsional (63.44%). Sementara bagi ViT, strategi ini bertindak sebagai prasyarat mutlak sekaligus katalisator yang menyembuhkan sifat *data hungry* nya (dari kondisi *underfitting* 31.11%), melepaskan potensi penuh arsitektur *transformer*, dan mendorong akurasi secara eksponensial hingga mendekati sempurna tanpa mengalami *overfitting*. Meskipun penelitian ini telah berhasil membuktikan superioritas arsitektur *Vision Transformer* dengan pencapaian akurasi yang sangat tinggi (97.40%), terdapat potensi pengembangan lanjutan yang dapat dieksplorasi oleh peneliti berikutnya: Optimasi dan Implementasi Model: Mengkaji teknik kompresi model (model *quantization* atau *pruning*) agar model ViT yang telah terlatih dengan sangat baik ini dapat diimplementasikan ke dalam aplikasi perangkat seluler (*mobile deployment*), sehingga manfaat pengenalan motif otomatis ini dapat langsung diakses oleh masyarakat dan UMKM batik. Ekspansi Skala Kelas Motif: Mengujian ketangguhan arsitektur ViT beserta strategi augmentasi dua tahap ini pada skala dataset yang lebih masif, misalnya dengan memperluas identifikasi motif batik Nusantara di luar 60 kelas Batik Nitik, guna menguji batas maksimal generalisasi model pada skala nasional.

DAFTAR PUSTAKA

- [1] Muhammad Rizki Pramudya Idris, Dzulfikar Azhar Al Ghifari, Gema Parasti Mindara, and Endang Purnama Giri, "Klasifikasi Motif Batik Indonesia Menggunakan Deep Learning Berbasis Convolutional Neural Network (CNN)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 10, no. 1, pp. 1140–1147, 2026, doi: 10.36040/jati.v10i1.16917.
- [2] Wahyuda Amurrohman and Nur Nafi'iyah, "PERBANDINGAN RESNET50, MOBILENETV2, DAN MOBILENETV3-LARGE DALAM KLASIFIKASI MOTIF BATIK INDONESIA BERBASIS TRANSFER LEARNING," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 11, no. 1, pp. 1457–1467, Jan. 2026, doi: 10.36341/rabit.v11i1.7372.
- [3] Ananda Rizki Dani and Irma Handayani, "Klasifikasi Motif Batik Yogyakarta Menggunakan Metode GLCM dan CNN," *Jurnal Teknologi Terpadu*, vol. 10, no. 2, pp. 142–156, 2024, doi: 10.54914/jtt.v10i2.1451.
- [4] Jeremia Daud Halim and Risal, "Klasifikasi Sel Darah Putih Menggunakan Vision Transformer (ViT)," *Jurnal Strategi*, vol. 6, no. 2, pp. 291–307, 2024, doi: 10.28932/strategi.v6i2.7303.
- [5] Fitrandi Ramadhana, Gayuh Abdi Mahardika, Nakata Day, and Eva Yulia Puspaningrum, "Klasifikasi Penyakit Daun Cabai Menggunakan Vision Transformer," *Seminar Nasional Informatika Bela Negara (SANTIKA)*, vol. 5, no. 1, pp. 131–135, 2025, doi: 10.33005/santika.v5i1.751.
- [6] Muhammad Mauladi and Dedy Hermanto, "Klasifikasi Motif Kain Jumpitan Palembang Menggunakan Metode CNN dengan Arsitektur Resnet-50," *Arcitech: Journal of Computer Science and Artificial Intelligence*, vol. 5, no. 2, pp. 332–355, Dec. 2025, doi: 10.29240/arcitech.v5i2.15310.
- [7] Aisyah Nur Salsabila, Muhaza Liebenlito, and Dhea Urfina Zulkifli, "Perbandingan Deteksi Alzheimer: ViT, CNN dan ViT dengan Bobot pada Citra Medis," *Indonesian Journal of Computer Science Attribution*, vol. 13, no. 1, pp. 1401–1412, 2024, doi: 10.33022/ijcs.v13i1.3765.
- [8] Akmal Hisyam Pradhana and Erna Daniati, "Benchmarking Vision Transformer Klasifikasi Visual Masakan Padang dengan ROBUSTNESS Melalui Augmentasi Data," *Seminar Nasional Teknologi dan Sistem Informasi (SITASI)*, vol. 5, no. 1, pp. 152–164, 2025, doi: 10.33005/sitasi.v5i1.2527.
- [9] Rachmat Santoso, "Augmentasi Data pada Prasasti Logam untuk Deteksi Aksara Kawi," *JURNAL FASILKOM*, vol. 14, no. 1, pp. 234–241, 2024, doi: 10.37859/jf.v14i1.6952.
- [10] Stefanus Eko Prasetyo, Haeruddin, and Wilsen Lau, "Analisis Komparatif Arsitektur CNN dan ViT dengan Variasi Augmentasi untuk Klasifikasi Karakter Mandarin," *BITNET Jurnal Pendidikan Teknologi Informasi*, vol. 11, no. 1, pp. 73–86, 2025, doi: 10.33084/bitnet.v11i1.11930.
- [11] Muhammad Ridwan Wibisono, Imam Ma'ruf Nugroho, and Meriska Defriani, "Batik Motif Classification Using CNN With Resnet-50 Architecture," *RISTEC: Research in Information Systems and Technology*, vol. 6, no. 1, pp. 41–59, 2025, doi: 10.35957/ristec.v6i1.1005.
- [12] Rizki Purnomo Pratama and Donny Avianto, "Implementasi Sistem Klasifikasi Batik Menggunakan MobileNet dengan Integrasi Chatbot Retrieval Augmented Generation," *BULLETIN OF COMPUTER SCIENCE RESEARCH*, vol. 6, no. 1, pp. 75–84, 2025, doi: 10.47065/bulletincsr.v6i1.863.
- [13] Arya Pangestu, Bedy Purnama, and Risnandar, "Vision Transformer untuk Klasifikasi Kematangan Pisang," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 75–84, Feb. 2024, doi: 10.25126/jtiik.20241117389.
- [14] Muhammad Fakhri Fadhlurrahman, Munir, and Yaya Wihardi, "Pengenalan Ekspresi Wajah Peserta Didik di Ruang Kelas Menggunakan Vision Transformer (ViT)," *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, vol. 4, no. 2, pp. 1047–1058, Aug. 2025, doi: 10.62712/juktisi.v4i2.531.
- [15] Stella Juventia Grace and Dedi Gunawan, "Perbandingan CNN, Resnet50, dan Vision

- Transformer untuk Klasifikasi Kanker Payudara Berbasis Web,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 945–956, Jul. 2025, doi: 10.36341/rabit.v10i2.6420.
- [16] Joseph Sanjaya and Mewati Ayub, “Augmentasi Data Pengenalan Citra Mobil Menggunakan Pendekatan Random Crop, Rotate, dan Mixup,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, pp. 311–323, Aug. 2020, doi: 10.28932/jutisi.v6i2.2688.
- [17] Desy Putri Ayuni, Jasril, Muhammad Irsyad, Febi Yanto, and Suwanto Sanjaya, “Augmentasi Data Pada Implementasi Convolutional Neural Network Arsitektur EFFICIENTNET-B3 untuk Klasifikasi Penyakit Daun Padi,” *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 2, pp. 239–249, 2023, doi: 10.31849/zn.v5i2.13874.
- [18] Ibnu Topan Adib Amrulloh, Betha Nurina Sari, and Tesa Nur Padilah, “Evaluasi Augmentasi Data Pada Deteksi Penyakit Daun Tebu dengan YOLOV8,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 7547–7552, 2024, doi: 10.36040/jati.v8i4.10267.
- [19] Riestiya Zain Fadillah, Ade Irawan, and Meredita Susanty, “Data Augmentasi Untuk Mengatasi Keterbatasan Data Pada Model Penerjemah Bahasa Isyarat Indonesia (BISINDO),” *JURNAL INFORMATIKA*, vol. 8, no. 2, pp. 208–214, 2021, doi: 10.31294/inf.v8i2.10768.
- [20] Rizki Nurhaliza Harahap and Kemas Muslim, “Peningkatan Akurasi pada Prediksi Kepribadian MbtI Pengguna Twitter Menggunakan Augmentasi Data,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 4, pp. 815–822, 2020, doi: 10.25126/jtiik.202073622.
- [21] Siti Auliaddina and Toni Arifin, “Penggunaan Data Augmentasi dan Hyperparameter Tuning dalam Klasifikasi Jenis Batik menggunakan Model CNN Use of Augmentation Data and Hyperparameter Tuning in Batik Type Classification Using the CNN Model,” *SISTEMASI*, vol. 13, no. 1, pp. 2540–9719, 2024, doi: 10.32520/stmsi.v13i1.3395