

ANALISIS PENGARUH STRATEGI AUGMENTASI DATA TERHADAP PERFORMA MODEL HYBRID CNN–SWIN TRANSFORMER DALAM KLASIFIKASI CITRA MIKROSKOPIS MALARIA

Muhammad Nabil Ramadhani¹, M. Muharrom Al Haromainy², Henni Endah Wahanani³

^{1, 2, 3}Universitas Pembangunan Nasional “Veteran” Jawa Timur, Jl. Rungkut Madya, Jawa Timur, Indonesia

Email: 22081010245@student.upnjatim.ac.id

Article History

Received: 02-07-2026

Revision: 02-08-2026

Accepted: 07-08-2026

Published: 10-08-2026

Abstract. Malaria diagnosis through microscopic images still faces challenges due to variations in the shape, size, and quality of blood cell images that can reduce the generalization ability of deep learning models. One approach to address this problem is data augmentation, but the effectiveness of various augmentation strategies on a hybrid Convolutional Neural Network (CNN) and Swin Transformer model has not been systematically compared. This study aims to compare the effect of five data augmentation strategies on the performance of the Hybrid CNN–Swin Transformer model in classifying malaria microscopic images. The NIH Malaria dataset consisting of 27,558 blood cell images was divided using the stratified split method with a ratio of 80:10:10. The model was trained using five augmentation strategies, namely Baseline (A0), Basic (A1), Geometric (A2), Advanced (A3), and RandAugment (A4), then evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and Train-Val Gap. All strategies produced accuracy above 96% and ROC-AUC above 99%. Strategy A3 provided the highest classification performance with an accuracy of 97.79%, an F1-score of 97.78%, and an ROC-AUC of 99.52%, while strategy A2 showed the most stable generalization ability based on the Train-Val Gap value. The results of this study provide empirical evidence that the choice of augmentation strategy affects the performance and generalization of the model, and serve as a reference in the development of a Hybrid CNN–Swin Transformer-based malaria classification system.

Keywords: Accuracy, Data Augmentation, Malaria Classification, Swin Transformer, Transfer Learning.

Abstrak. Diagnosis malaria melalui citra mikroskopis masih menghadapi tantangan karena variasi bentuk, ukuran, dan kualitas citra sel darah yang dapat menurunkan kemampuan generalisasi model *deep learning*. Salah satu pendekatan untuk mengatasi masalah tersebut adalah augmentasi data, namun efektivitas berbagai strategi augmentasi pada model hibrida *Convolutional Neural Network* (CNN) dan *Swin Transformer* masih belum banyak dibandingkan secara sistematis. Penelitian ini bertujuan membandingkan pengaruh lima strategi augmentasi data terhadap performa model *Hybrid CNN–Swin Transformer* dalam klasifikasi citra mikroskopis malaria. Dataset NIH Malaria yang terdiri atas 27.558 citra sel darah dibagi menggunakan metode *stratified split* dengan rasio 80:10:10. Model dilatih menggunakan lima strategi augmentasi, yaitu *Baseline* (A0), *Basic* (A1), *Geometric* (A2), *Advanced* (A3), dan *RandAugment* (A4), kemudian dievaluasi menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *ROC-AUC*, dan *Train-Val Gap*. Seluruh strategi menghasilkan *accuracy* di atas 96% dan *ROC-AUC* di atas 99%. Strategi A3 memberikan performa klasifikasi tertinggi dengan *accuracy* 97,79%, *F1-score* 97,78%, dan *ROC-AUC* 99,52%, sedangkan strategi A2 menunjukkan kemampuan generalisasi paling stabil berdasarkan nilai *Train-Val Gap*. Hasil penelitian ini memberikan bukti empiris bahwa pemilihan strategi augmentasi berpengaruh terhadap performa dan generalisasi model, serta menjadi acuan dalam pengembangan sistem klasifikasi malaria berbasis *Hybrid CNN–Swin Transformer*.

Kata Kunci: Accuracy, Augmentasi Data, Klasifikasi Malaria, Swin Transformer, Transfer Learning

How to Cite: Ramadhani, M. N., Haromainy, M. M. A., & Wahanani, H. E. (2026). Analisis Pengaruh Strategi Augmentasi Data Terhadap Performa Model Hybrid Cnn–Swin Transformer dalam Klasifikasi Citra Mikroskopis Malaria. *HORIZON: Indonesian Journal of Multidisciplinary*, 4 (4), 4740-4751. <http://doi.org/10.54373/hijm.v4i4.7178>

PENDAHULUAN

Malaria merupakan penyakit infeksi yang disebabkan oleh parasit *Plasmodium* dan ditularkan melalui gigitan nyamuk *Anopheles* betina. Penyakit ini masih menjadi salah satu masalah kesehatan utama di berbagai negara tropis dan subtropis. Menurut *World Health Organization* (WHO), malaria masih menyebabkan ratusan juta kasus setiap tahun, sedangkan di Indonesia sebagian besar kasus terkonsentrasi di wilayah Papua. Tingginya angka kejadian tersebut menunjukkan bahwa diagnosis yang cepat dan akurat sangat diperlukan untuk mendukung penanganan pasien serta menurunkan angka kesakitan dan kematian akibat malaria (WHO, 2023).

Pemeriksaan mikroskopis preparat darah masih menjadi standar emas dalam diagnosis malaria. Namun, metode ini sangat bergantung pada pengalaman analis laboratorium, memerlukan waktu pemeriksaan yang relatif lama, serta berpotensi menghasilkan kesalahan identifikasi, terutama pada preparat dengan kualitas citra rendah atau tingkat parasitemia yang kecil (Masud et al., 2020; Nurhaeni et al., 2024). Kondisi tersebut mendorong berkembangnya pemanfaatan *deep learning* untuk mengotomatisasi proses klasifikasi citra mikroskopis sehingga diagnosis dapat dilakukan secara lebih objektif, konsisten, dan efisien.

Berbagai penelitian telah menunjukkan bahwa *Convolutional Neural Network* (CNN) mampu memberikan akurasi tinggi dalam klasifikasi citra malaria karena efektif mengekstraksi fitur lokal dari citra (Pan et al., 2018; Huda et al., 2022; Saksenata et al., 2022; Hasyani et al., 2023; Purnomo et al., 2024). Namun, CNN memiliki keterbatasan dalam memodelkan hubungan spasial global akibat penggunaan kernel konvolusi dengan *receptive field* yang terbatas. Sebaliknya, *Vision Transformer* (ViT), khususnya *Swin Transformer*, memanfaatkan mekanisme *self-attention* dan *shifted window* untuk menangkap hubungan global antarfitur secara lebih efektif sehingga mampu menghasilkan representasi citra yang lebih kaya (Dosovitskiy et al., 2021; Liu et al., 2021; Chen et al., 2023). Oleh karena itu, penggabungan CNN dan *Swin Transformer* dalam arsitektur *Hybrid CNN–Swin Transformer* menjadi pendekatan yang menjanjikan karena mampu menggabungkan keunggulan ekstraksi fitur lokal dan pemodelan konteks global (Xiao et al., 2021; Khan et al., 2023; Long, 2024; Oliveira et al., 2024; Kim et al., 2025).

Selain pemilihan arsitektur, performa model *deep learning* juga sangat dipengaruhi oleh strategi augmentasi data yang digunakan selama proses pelatihan. Augmentasi data bertujuan meningkatkan keragaman citra sehingga model lebih mampu melakukan generalisasi dan mengurangi risiko *overfitting*. Berbagai teknik telah dikembangkan, mulai dari transformasi dasar seperti rotasi dan *flip*, transformasi geometris, hingga pendekatan lanjutan seperti *MixUp*,

CutMix, dan *RandAugment*. Masing-masing strategi menghasilkan karakteristik data latih yang berbeda sehingga berpotensi memberikan pengaruh yang berbeda pula terhadap performa model.

Penelitian terdahulu umumnya berfokus pada dua aspek yang terpisah. Sebagian besar penelitian membandingkan berbagai arsitektur CNN untuk klasifikasi malaria tanpa mengevaluasi pengaruh strategi augmentasi yang digunakan (Huda et al., 2022; Hasyani et al., 2023; Purnomo et al., 2024; Setiawan, 2025). Sebaliknya, penelitian mengenai *Vision Transformer* dan model hibrida lebih banyak diarahkan pada pengembangan arsitektur untuk berbagai aplikasi pencitraan medis, tetapi belum mengkaji secara sistematis bagaimana variasi strategi augmentasi memengaruhi performa model hibrida tersebut (Long, 2024; Oliveira et al., 2024; Kim et al., 2025). Dengan demikian, belum terdapat kajian yang membandingkan secara komprehensif efektivitas berbagai strategi augmentasi, mulai dari tanpa augmentasi, augmentasi dasar, augmentasi geometris, augmentasi berbasis *MixUp–CutMix*, hingga *RandAugment*, pada model *Hybrid CNN–Swin Transformer* untuk klasifikasi citra mikroskopis malaria menggunakan protokol evaluasi yang sama. Kesenjangan ini menyebabkan belum tersedianya bukti empiris mengenai strategi augmentasi yang paling sesuai untuk meningkatkan performa sekaligus kemampuan generalisasi model hibrida.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan membandingkan pengaruh lima strategi augmentasi data terhadap performa model *Hybrid CNN–Swin Transformer* dalam klasifikasi citra mikroskopis malaria menggunakan NIH Malaria Dataset. Evaluasi dilakukan berdasarkan *accuracy*, *precision*, *recall*, *F1-score*, *ROC-AUC*, dan *Train-Val Gap* sebagai indikator kemampuan generalisasi model. Kebaruan penelitian ini terletak pada evaluasi komparatif lima strategi augmentasi yang diterapkan secara konsisten pada satu arsitektur *Hybrid CNN–Swin Transformer*, sehingga pengaruh masing-masing strategi dapat dianalisis secara objektif tanpa dipengaruhi perbedaan arsitektur model. Hasil penelitian ini diharapkan memberikan rekomendasi empiris mengenai strategi augmentasi yang paling efektif untuk meningkatkan akurasi dan generalisasi model pada klasifikasi citra mikroskopis malaria serta menjadi acuan bagi pengembangan sistem diagnosis malaria berbasis *deep learning*.

METODE

Dataset Penelitian

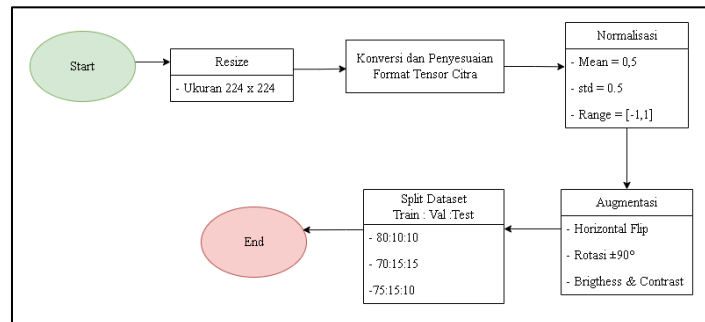
Bahan utama yang digunakan dalam penelitian ini adalah NIH Malaria Dataset yang dikembangkan oleh *National Library of Medicine* (NLM), *National Institutes of Health* (NIH), dan diperoleh melalui platform *Kaggle* (<https://www.kaggle.com/datasets/iarunava/cell->

[images-for-detecting-malaria](#)). Dataset ini terdiri dari 27.558 citra mikroskopis sel darah merah berformat PNG dengan resolusi bervariasi, terbagi secara seimbang ke dalam dua kelas, yaitu 13.779 citra sel darah terinfeksi (*infected*) dan 13.779 citra sel darah normal (*uninfected*). Citra diperoleh dari preparat darah tipis yang diamati menggunakan mikroskop cahaya dengan pewarnaan Giemsa. Morfologi eritrosit dan karakter visual parasit Plasmodium yang relatif universal menjadikan dataset ini representatif untuk penelitian klasifikasi malaria berbasis citra.

Tabel 1. Distribusi Dataset NIH Malaria

No	Kelas	Jumlah Citra	Sumber
1	<i>Infected</i> (Terinfeksi)	13.779	NIH / Kaggle
2	<i>Uninfected</i> (Normal)	13.779	NIH / Kaggle
Total		27.558	

Preprocessing Data



Gambar 1. Alur *Preprocessing* Data

Seluruh citra input diproses melalui empat tahap *preprocessing* sebelum dimasukkan ke dalam model. Pertama, *resize*: seluruh citra diseragamkan menjadi ukuran 224×224 piksel agar sesuai dengan kebutuhan input model *Hybrid CNN–Swin Transformer* dan meningkatkan efisiensi komputasi. Kedua, konversi tensor: nilai piksel diubah dari rentang integer 0–255 menjadi floating point 0–1 dengan susunan dimensi *Channel × Height × Width* ($C \times H \times W$) agar kompatibel dengan lapisan konvolusi CNN maupun modul Transformer. Ketiga, normalisasi: diterapkan menggunakan nilai mean 0,5 dan standar deviasi 0,5 untuk menggeser rentang nilai piksel menjadi -1 hingga 1 sehingga distribusi data terpusat di sekitar nol dan pelatihan model berlangsung lebih stabil. Keempat, augmentasi data: diterapkan secara *on-the-fly* pada data latih di setiap *epoch*, sementara data validasi dan data uji tidak dikenai augmentasi agar evaluasi tetap objektif.

Pembagian Dataset

Pembagian dataset dilakukan menggunakan teknik *stratified split* untuk memastikan proporsi kelas *infected* dan *uninfected* tetap seimbang pada setiap subset. Penelitian ini menggunakan rasio pembagian 80:10:10, menghasilkan 22.046 citra data latih, 2.756 citra data validasi, dan 2.756 citra data uji.

Tabel 2. Pembagian Dataset

No	Rasio Split	Data Train	Data Validasi	Data Test
1	80:10:10	22.046	2.756	2.756

Konfigurasi Pelatihan

Seluruh eksperimen dijalankan dengan konfigurasi *hyperparameter* yang seragam untuk memastikan perbandingan yang adil antara strategi augmentasi. Konfigurasi yang digunakan meliputi *batch size* 16, *learning rate* 1×10^{-4} , dan *weight decay* 1×10^{-2} . *Optimizer* yang digunakan adalah *AdamW* dengan penjadwalan *learning rate* menggunakan *cosine annealing*. Pelatihan dilakukan selama jumlah *epoch* yang ditetapkan dengan *early stopping* berdasarkan *validation loss* untuk mencegah *overfitting*.

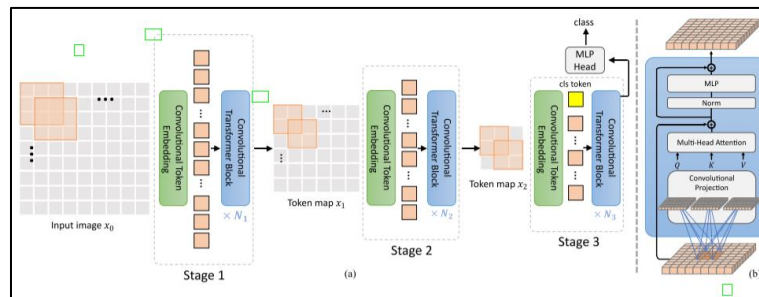
Metrik Evaluasi

Performa model dievaluasi menggunakan lima metrik kuantitatif. *Accuracy* mengukur proporsi prediksi yang benar secara keseluruhan. *Precision* mengukur ketepatan model dalam memprediksi kelas positif (*infected*). *Recall* atau sensitivitas mengukur kemampuan model mendeteksi seluruh kasus positif yang sebenarnya metrik ini sangat kritis dalam konteks klinis karena *false negative* berpotensi menyebabkan pasien terinfeksi tidak tertangani. *F1-score* merupakan rata-rata harmonik dari *precision* dan *recall* yang memberikan gambaran keseimbangan antara keduanya. *ROC-AUC* mengukur kemampuan diskriminasi model antara kelas *infected* dan *uninfected* secara keseluruhan terlepas dari *threshold* yang dipilih. Sebagai metrik tambahan, *Train-Val Gap* (selisih *accuracy* data latih dan validasi) digunakan sebagai indikator kuantitatif *overfitting*, di mana nilai mendekati nol mengindikasikan generalisasi yang optimal.

Arsitektur Model

Model yang digunakan adalah arsitektur Hybrid CNN–Swin Transformer yang mengintegrasikan kemampuan ekstraksi fitur lokal CNN dengan pemodelan konteks global Swin Transformer. Arsitektur ini terdiri dari tiga komponen utama. Pertama, Blok CNN: tiga lapisan konvolusi berukuran kernel 3×3 dengan fungsi aktivasi ReLU dan max pooling 2×2

bertugas mengekstraksi fitur lokal dari citra input. Kedua, Bridge Layer: berfungsi sebagai titik integrasi yang mengonversi *feature map* hasil CNN menjadi *token sequence* melalui proses *patch partition* dan *linear embedding* berdimensi 96, sehingga representasi fitur dapat diproses oleh modul Transformer. Ketiga, Swin Transformer Block: memproses *token sequence* menggunakan mekanisme Window-based Multi-Head Self-attention (W-MHSA) dan Shifted Window Multi-Head Self-Attention (SW-MHSA) untuk menangkap konteks global citra secara hierarkis. Pada tahap akhir, representasi fitur global diproses melalui *Global Average Pooling* (GAP) dan lapisan *fully connected* dengan fungsi aktivasi *softmax* untuk menghasilkan probabilitas klasifikasi biner, dengan *threshold* keputusan sebesar 0,5.



Gambar 2. Model Arsitektur Hybrid CNN–Swin Transformer

Definisi Strategi Augmentasi

Penelitian ini mendefinisikan lima strategi augmentasi sebagai variabel independen yang dibandingkan secara sistematis. Strategi A0 (*Baseline*) hanya menerapkan *resize* dan normalisasi tanpa augmentasi apapun, berfungsi sebagai pembanding dasar. Strategi A1 (*Basic*) menerapkan *horizontal flip*, rotasi 90°, dan *ColorJitter*, setara dengan augmentasi standar yang umum digunakan pada penelitian sebelumnya. Strategi A2 (*Geometric*) memperluas A1 dengan menambahkan *vertical flip*, *random crop*, dan *affine transform* untuk memperkaya variasi spasial secara lebih agresif. Strategi A3 (*Advanced/MixUp+CutMix*) menggabungkan A1 dengan teknik interpolasi citra *MixUp* ($\alpha=0,4$) dan *CutMix* ($\alpha=1,0$) yang diterapkan pada level batch selama *training loop*. Strategi A4 (*RandAugment*) menerapkan augmentasi berbasis kebijakan otomatis dengan parameter $N=2$ (jumlah transformasi per sampel) dan $M=9$ (magnitudo transformasi) yang dipilih secara adaptif tanpa tuning manual.

Tabel 3. Strategi Augmentasi

Kode	Nama Strategi	Teknik Augmentasi	Keterangan
A0	<i>Baseline</i>	<i>Resize</i> + Normalisasi saja	Tanpa augmentasi, sebagai pembanding dasar
A1	<i>Basic</i>	<i>Horizontal flip</i> + Rotasi 90° + <i>ColorJitter</i>	Sama dengan augmentasi pada skripsi acuan

A2	<i>Geometric</i>	A1 + <i>Vertical flip</i> + <i>Random crop</i> + <i>Affine transform</i>	Memperkaya variasi spasial lebih agresif
A3	<i>Advanced</i>	A1 + <i>MixUp</i> ($\alpha=0.4$) + <i>CutMix</i> ($\alpha=1.0$)	Diterapkan pada batch level saat <i>training loop</i>

HASIL DAN DISKUSI

Pengaruh Strategi Augmentasi terhadap Performa Model Hybrid CNN–Swin Transformer

Tabel 4. Perbandingan Performa Model Hybrid CNN–Swin Transformer Berdasarkan Strategi Augmentasi Data

Strategi	Nama	Acc	Prec	Rec	F1	AUC
A0	<i>Baseline</i> (No Aug)	0.9692	0.9675	0.9710	0.9692	0.9940
A1	<i>Basic</i> (Flip+Rotate+CJ)	0.9684	0.9708	0.9659	0.9684	0.9944
A2	<i>Geometric</i> (A1+VFlip+Crop+Affine)	0.9750	0.9788	0.9710	0.9749	0.9952
A3	<i>Advanced</i> (A1+MixUp+CutMix)	0.9779	0.9810	0.9746	0.9778	0.9956
A4	<i>RandAugment</i> (N=2, M=9)	0.9746	0.9795	0.9695	0.9745	0.9951

Hasil eksperimen menunjukkan bahwa strategi augmentasi data memberikan pengaruh terhadap kemampuan klasifikasi model *Hybrid CNN–Swin Transformer*. Berdasarkan Tabel 4, strategi A3 (*Advanced: MixUp + CutMix*) menghasilkan performa terbaik dengan *accuracy* 97,79%, *precision* 98,10%, *recall* 97,46%, *F1-score* 97,78%, dan *ROC-AUC* 99,56%, sedangkan strategi A1 (*Basic*) memberikan performa terendah. Temuan ini menunjukkan bahwa augmentasi berbasis interpolasi sampel lebih efektif dalam meningkatkan kemampuan generalisasi model dibandingkan augmentasi konvensional yang hanya memodifikasi orientasi atau posisi citra.

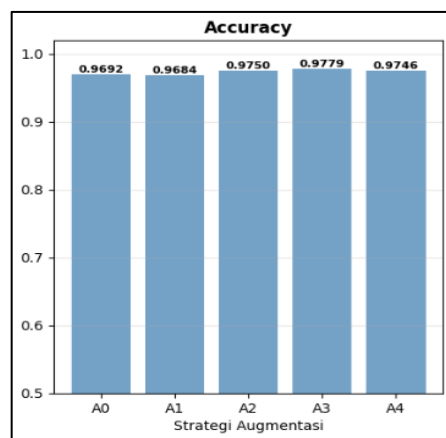
Keunggulan *MixUp* dan *CutMix* dapat dijelaskan melalui karakteristik citra mikroskopis malaria. Pada preparat darah, sel darah merah memiliki bentuk yang relatif seragam, sedangkan perbedaan antara sel *infected* dan *uninfected* sering kali hanya ditunjukkan oleh keberadaan parasit berukuran kecil, perubahan tekstur sitoplasma, atau variasi intensitas pewarnaan. Selain itu, kualitas citra mikroskopis dapat dipengaruhi oleh variasi pencahayaan, fokus mikroskop, artefak preparat, dan tingkat pewarnaan yang berbeda. Kondisi tersebut menyebabkan model berpotensi mempelajari karakteristik visual yang bersifat spesifik terhadap data latih (*overfitting*), bukan pola biologis yang benar-benar merepresentasikan infeksi malaria. *MixUp* mengatasi permasalahan tersebut dengan membentuk sampel baru melalui interpolasi dua citra beserta labelnya, sehingga model terdorong mempelajari batas keputusan (*decision boundary*) yang lebih halus dan tidak terlalu sensitif terhadap variasi individual setiap citra. Sementara

itu, *CutMix* melatih model mengenali keberadaan fitur infeksi meskipun hanya muncul pada sebagian area citra, sehingga model tidak bergantung pada satu lokasi atau pola visual tertentu.

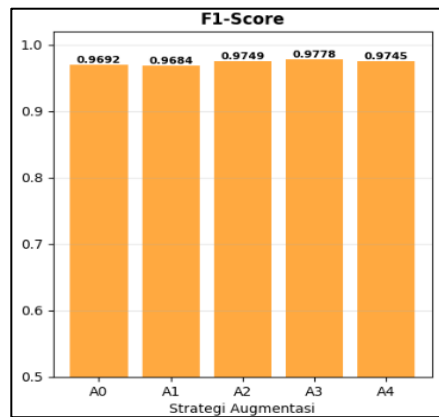
Karakteristik tersebut juga sesuai dengan arsitektur *Hybrid CNN–Swin Transformer*. CNN berperan mengekstraksi fitur lokal, seperti tekstur eritrosit, bentuk inti parasit, dan perubahan morfologi sel, sedangkan *Swin Transformer* memanfaatkan mekanisme *self-attention* untuk mempelajari hubungan antarpatch citra secara global (Liu et al., 2021; Chen et al., 2023). Variasi citra yang dihasilkan oleh *MixUp* dan *CutMix* membuat model belajar mengintegrasikan informasi lokal dan global secara lebih seimbang, sehingga representasi fitur menjadi lebih robust terhadap perubahan posisi parasit, variasi pewarnaan, maupun kualitas preparat. Hal ini menjelaskan mengapa strategi A3 menghasilkan performa klasifikasi yang lebih tinggi dibandingkan augmentasi geometri yang umumnya hanya menghasilkan transformasi spasial tanpa memperkaya distribusi fitur.

Temuan penelitian ini sejalan dengan Khan et al. (2023), yang menyatakan bahwa augmentasi lanjutan mampu meningkatkan kemampuan generalisasi model berbasis *Transformer* melalui pembentukan representasi fitur yang lebih robust. Hasil ini juga mendukung ulasan Long (2024) dan Kim et al. (2025), yang menunjukkan bahwa model hibrida *CNN-Transformer* memperoleh manfaat lebih besar dari augmentasi kompleks dibandingkan CNN konvensional karena mampu memanfaatkan informasi lokal dan global secara bersamaan. Secara khusus pada klasifikasi citra mikroskopis malaria, penelitian ini memberikan bukti empiris bahwa strategi *MixUp* dan *CutMix* tidak hanya meningkatkan performa klasifikasi, tetapi juga lebih efektif dalam mengatasi variasi morfologi sel darah dan kualitas citra mikroskopis yang menjadi tantangan utama dalam pengembangan sistem diagnosis malaria berbasis *deep learning*.

Analisis Performa Accuracy dan F1-score



Gambar 3. Hasil Evaluasi Augmentasi per Metrik Accuracy



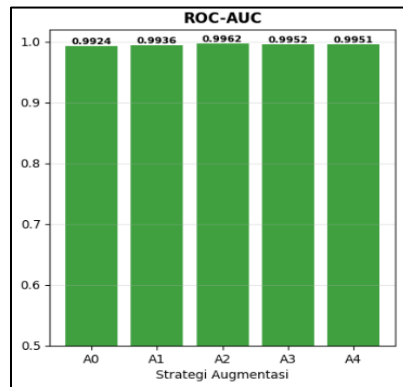
Gambar 4. Hasil Evaluasi Augmentasi per Metrik *F1-score*

Berdasarkan gambar diatas, nilai *accuracy* dan *F1-score* menunjukkan pola peningkatan yang relatif konsisten pada setiap strategi augmentasi. Strategi *Advanced* (A3) menghasilkan performa terbaik dengan *accuracy* sebesar 97,79% dan *F1-score* sebesar 97,78%, sedangkan strategi *Basic* (A1) menghasilkan performa terendah dengan *accuracy* 96,84%. Hasil tersebut menunjukkan bahwa penerapan augmentasi tingkat lanjut mampu meningkatkan kemampuan model dalam mengenali kedua kelas secara lebih seimbang dibandingkan augmentasi dasar maupun tanpa augmentasi. Selain menghasilkan *accuracy* tertinggi, strategi A3 juga memperoleh nilai *recall* sebesar 97,46%, yang mengindikasikan kemampuan model dalam mendeteksi sebagian besar citra sel darah yang terinfeksi. Dalam konteks diagnosis malaria, peningkatan *recall* sangat penting karena dapat mengurangi kemungkinan terjadinya *false negative*, yaitu kondisi ketika sel darah yang terinfeksi diklasifikasikan sebagai normal. Peningkatan tersebut menunjukkan bahwa kombinasi *MixUp* dan *CutMix* mampu memperkaya variasi data pelatihan sehingga model menjadi lebih robust terhadap perubahan bentuk parasit, intensitas pewarnaan Giemsa, maupun variasi morfologi eritrosit. Hasil penelitian ini sejalan dengan penelitian Chen et al. (2023) yang menyatakan bahwa arsitektur *Swin Transformer* memiliki kemampuan lebih baik dalam menangkap variasi visual pada citra sel darah dibandingkan CNN konvensional.

Analisis ROC-AUC dan Kemampuan Diskriminasi Model

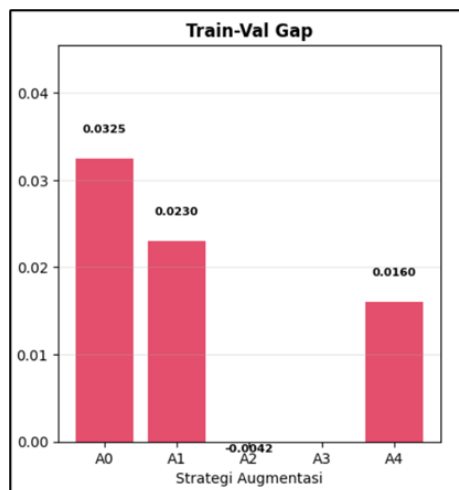
Seluruh strategi augmentasi menghasilkan nilai *ROC-AUC* di atas 99%, yang menunjukkan kemampuan diskriminasi model yang sangat baik dalam membedakan kelas *infected* dan *uninfected*. Strategi *Advanced* (A3) memperoleh nilai *ROC-AUC* tertinggi sebesar 99,56%, sedangkan *Baseline* (A0) memperoleh 99,40%. Meskipun perbedaan *accuracy* antar strategi relatif kecil, peningkatan nilai *ROC-AUC* menunjukkan bahwa augmentasi tingkat

lanjut mampu menghasilkan representasi fitur yang lebih baik. Hasil ini mengindikasikan bahwa kombinasi CNN sebagai pengekstraksi fitur lokal dan *Swin Transformer* sebagai pemodel konteks global melalui mekanisme *self-attention* mampu meningkatkan kemampuan klasifikasi citra.



Gambar 5. Pengaruh Data Split Pada Metrik Evaluasi

Pengaruh Strategi Augmentasi terhadap *Overfitting*



Gambar 6. Ranking Skenario *Hyperparameter*

Analisis *Train-Val Gap* menunjukkan bahwa strategi augmentasi data berperan dalam mengurangi *overfitting*. Strategi *Baseline* (A0) menghasilkan *gap* tertinggi sebesar 0,0325, sedangkan strategi *Geometric* (A2) dan *Advanced* (A3) menghasilkan *gap* yang lebih kecil sehingga menunjukkan kemampuan generalisasi yang lebih baik. Strategi A2 bahkan menghasilkan *Train-Val Gap* negatif ($-0,0042$), yang menunjukkan bahwa performa pada data validasi sedikit lebih tinggi daripada data pelatihan. Kondisi ini mengindikasikan bahwa augmentasi menghasilkan variasi data yang cukup beragam sehingga model tidak hanya menghafal data pelatihan, tetapi mampu membangun representasi fitur yang lebih robust dan adaptif terhadap data baru.

KESIMPULAN

Hasil penelitian menunjukkan bahwa strategi augmentasi data memberikan pengaruh terhadap performa model Hybrid CNN–Swin Transformer dalam klasifikasi citra mikroskopis malaria. Di antara lima strategi yang diuji, strategi Advanced (A3) berbasis MixUp dan CutMix menghasilkan performa paling baik, sedangkan strategi Geometric (A2) menunjukkan kemampuan generalisasi yang paling stabil berdasarkan nilai Train-Val Gap. Temuan ini mengindikasikan bahwa penerapan augmentasi yang lebih kompleks mampu meningkatkan kualitas representasi fitur sekaligus mengurangi kecenderungan overfitting selama proses pelatihan.

Penelitian ini menunjukkan bahwa pemilihan strategi augmentasi merupakan faktor penting yang mendukung keberhasilan model selain pemilihan arsitektur Hybrid CNN–Swin Transformer. Oleh karena itu, strategi augmentasi yang tepat dapat menjadi pertimbangan dalam pengembangan sistem klasifikasi citra medis berbasis deep learning. Penelitian selanjutnya dapat mengevaluasi strategi augmentasi lainnya pada dataset yang lebih beragam serta membandingkan performanya dengan arsitektur Vision Transformer yang lebih mutakhir untuk memperoleh hasil yang lebih komprehensif

REFERENSI

- Chen, S., Lu, S., Wang, S., Ni, Y., & Zhang, Y. (2023). Shifted window vision transformer for blood cell classification. *Electronics*, 12(11). <https://doi.org/10.3390/electronics12112442>
- Dosovitskiy, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. <http://arxiv.org/abs/2010.11929>
- Saksenata, A. F., Minarno, A. E., & Azhar, Y. (2022). Klasifikasi citra sel darah untuk penyakit malaria dengan metode CNN. *REPOSITOR*, 4(2), 185–194.
- Hasyani, R. A., Simbolon, S. M., Mufida, Y., Ester, Y., & Ritonga, B. (2023). Klasifikasi malaria melalui penggunaan convolutional neural network pada citra sel darah. *Student Research Journal*, 1(6). <https://doi.org/10.55606/srjyappi.v1i6.853>
- Huda, N., Prayogi, S. Y., Ahmad, M. A., & Dewi, A. Y. (2022). Klasifikasi malaria menggunakan metode image processing dari sel darah merah dengan algoritma convolutional neural network. *JOINS (Journal of Information System)*, 7(2), 166–177. <https://doi.org/10.33633/joins.v7i2.7068>
- Khan, A., et al. (2023). A survey of the vision transformers and their CNN-transformer based variants. *Artificial Intelligence Review*, 56, 2917–2970. <https://doi.org/10.1007/s10462-023-10595-0>
- Kim, J. W., Khan, A. U., & Banerjee, I. (2025). Systematic review of hybrid vision transformer architectures for radiological image analysis. *Journal of Digital Imaging*. <https://doi.org/10.1007/s10278-024-01322-4>
- Liu, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. <http://arxiv.org/abs/2103.14030>

- Long, H. (2024). Hybrid design of CNN and vision transformer: A review. In *Proceedings of the ACM International Conference* (pp. 121–127). Association for Computing Machinery. <https://doi.org/10.1145/3703187.3703208>
- Masud, M., et al. (2020). Leveraging deep learning techniques for malaria parasite detection using mobile application. *Wireless Communications and Mobile Computing*, 2020. <https://doi.org/10.1155/2020/8895429>
- Nurhaeni, N., Prastya, S. E., Hidayat, A., & Anisa, F. N. (2024). Pemodelan sistem deteksi parasit malaria pada citra mikroskopis sel darah menggunakan metode deep learning. *SMATIKA JURNAL*, 14(2), 409–416. <https://doi.org/10.32664/smatika.v14i02.1475>
- Oliveira, D. D. L. L., Tosta, T. A. A., Neves, L. A., & Nascimento, M. D. Z. (2024). Hybrid CNN-transformer models in histopathology image analysis: A scoping review. <https://doi.org/10.1109/ACCESS.2024.0429000>
- Pan, W. D., Dong, Y., & Wu, D. (2018). Classification of malaria-infected cells using deep convolutional neural networks. In *Machine Learning: Advanced Techniques and Emerging Applications*. InTech. <https://doi.org/10.5772/intechopen.72426>
- Purnomo, J. G., Birowo, S., & Maulana, M. A. (2024). Identifikasi malaria pada citra darah dengan convolutional neural network. *bit-Tech*, 7(2), 406–412. <https://doi.org/10.32877/bt.v7i2.1828>
- Sari, Z., Azhar, Y., & Yusuf, A. (2024). Deteksi penyakit malaria menggunakan klasifikasi berbasis CNN. *Syntax Literate*, 9(1). <https://doi.org/10.36418/syntax-literate.v9i1>
- Setiawan, A. W. (2025). Perbandingan kinerja arsitektur convolutional neural network pada deteksi malaria menggunakan citra sel darah (*Comparison of performance of convolutional neural networks architectures on malaria detection using blood cell image*), 11(3), 643–652.
- Xiao, T., Singh, M., Mintun, E., Darrell, T., Dollár, P., & Girshick, R. (2021). *Early convolutions help transformers see better*. <http://arxiv.org/abs/2106.14881>