

KLASIFIKASI PENYAKIT PNEUMONIA MENGGUNAKAN MODEL HYBRID CNN-TRANSFORMER BERBASIS CITRA X-RAY PARU-PARU

Nur Milani Hidayah¹⁾, Muhammad Faisal²⁾, Desi Anggreani³⁾, Nurnawaty⁴⁾, Andi Makbul Syamsuri⁵⁾, M. Agusalim⁶⁾

1. Teknik, Universitas Muhammadiyah Makassar
email: 105841100822@student.unismuh.ac.id
2. Teknik, Universitas Muhammadiyah Makassar
email: muhfaisal@unismuh.ac.id
3. Teknik, Universitas Muhammadiyah Makassar
email: desianggreani@unismuh.ac.id
4. Teknik Pengairan, Universitas Muhammadiyah Makassar
email: nurnawaty@unismuh.ac.id
5. Teknik Pengairan, Universitas Muhammadiyah Makassar
email: amakbulsyamsuri@unismuh.ac.id
6. Teknik Pengairan, Universitas Muhammadiyah Makassar
email: m.agusalim@unismuh.ac.id

Abstract

This study aims to apply a Hybrid CNN-Transformer model based on Medical Vision Transformer (MedViT) for pneumonia classification using chest X-Ray images. The dataset consisted of 450 images, including 150 pneumonia images, 150 non-pneumonia images, and 150 random images as a control class to test system robustness. The data were obtained from Labuang Baji Hospital, Makassar, during the 2023 to 2025 period. The research stages included data collection, preprocessing, augmentation, dataset splitting, model implementation, training, and performance evaluation. The tested models consisted of CNN, Vision Transformer (ViT), and Hybrid CNN-Transformer. The evaluation used accuracy, precision, recall, F1-score, AUC, confusion matrix, ROC curve, and Grad-CAM visualization. The results showed that the Hybrid CNN-Transformer model achieved the best performance with an accuracy of 95.59%, precision of 96.12%, recall of 95.59%, F1-score of 95.58%, and AUC of 0.9968. The model improved accuracy by 8.83% compared with CNN and produced fewer classification errors. The Grad-CAM visualization also indicated that the model focused on relevant lung areas. These findings indicate that combining CNN local feature extraction with Transformer global context can improve pneumonia classification based on medical images..

Kata Kunci : CNN; Transformer; Pneumonia; X-Ray; Deep Learning

A. PENDAHULUAN

Pneumonia merupakan penyakit infeksi pada saluran pernapasan bawah

yang menyerang jaringan paru-paru, khususnya *alveoli*. Penyakit ini dapat menyebabkan gangguan pertukaran

oksigen dan ditandai oleh batuk, demam, nyeri dada, serta kesulitan bernapas. Pada kelompok rentan, terutama anak-anak dan lanjut usia, *pneumonia* masih menjadi masalah kesehatan yang perlu mendapat perhatian serius (Zahra & Rosfadilla, 2025).

Kota Makassar sebagai salah satu pusat layanan kesehatan di Indonesia Timur menghadapi kebutuhan pemeriksaan penyakit pernapasan yang cukup tinggi. RS Labuang Baji menjadi salah satu fasilitas kesehatan yang menangani pasien dengan keluhan pernapasan dan menggunakan pemeriksaan radiografi dada atau *chest X-Ray*. Dalam praktik klinis, citra *X-Ray* membantu tenaga medis mengenali adanya opasitas, infiltrat, atau konsolidasi pada paru-paru (Taula'bi et al., 2024).

Interpretasi citra *X-Ray* tidak selalu mudah. Perbedaan antara citra paru *non pneumonia* dan *pneumonia* dapat tampak samar, terutama ketika pola infeksi tersebar halus pada beberapa area paru. Kondisi ini menuntut ketelitian tinggi dan waktu analisis yang lebih panjang. Karena itu, sistem berbasis komputasi dapat digunakan sebagai alat bantu untuk mempercepat proses klasifikasi dan meningkatkan konsistensi analisis citra medis (Rahmadewi et al., 2023).

Convolutional Neural Network (CNN) banyak digunakan dalam klasifikasi citra karena mampu mengekstraksi fitur lokal, seperti tepi, tekstur, dan pola intensitas. Namun, CNN memiliki keterbatasan dalam menangkap hubungan spasial jarak jauh karena proses konvolusi berfokus pada area lokal tertentu (Mathi et al., 2025). Sebaliknya, *Vision Transformer* (ViT) memanfaatkan mekanisme *self-attention* untuk memahami relasi global antarbagian citra, tetapi membutuhkan data yang cukup besar agar stabil saat pelatihan (Takahashi et al., 2024).

Pendekatan *Hybrid CNN-Transformer*, khususnya *Medical Vision Transformer* (MedViT), menggabungkan keunggulan CNN dan *Transformer*. CNN berperan menangkap fitur lokal, sedangkan *Transformer* memodelkan konteks global citra. Integrasi ini relevan untuk klasifikasi *pneumonia* karena pola kelainan paru dapat muncul sebagai perubahan kecil maupun sebaran luas pada citra *X-Ray* (Manzari et al., 2023). Berdasarkan masalah tersebut, penelitian ini menerapkan model *Hybrid CNN-Transformer* untuk klasifikasi *pneumonia* berbasis citra *X-Ray* paru-paru.

Beberapa penelitian terdahulu telah menerapkan metode *Convolutional Neural Network* (CNN) untuk deteksi *pneumonia* dan mencapai akurasi yang cukup baik. Namun, model konvolusional murni kerap mengalami keterbatasan dalam mengekstraksi hubungan spasial antar-region paru yang berjauhan. Di sisi lain, penggunaan *Vision Transformer* (ViT) secara independen membutuhkan pasokan data masif agar tidak rentan *overfitting*. Pendekatan hibrida melalui *Medical Vision Transformer* (MedViT) menawarkan solusi kompromistis dengan menggabungkan keunggulan ekstraksi fitur lokal CNN dan mekanisme *self-attention global Transformer*. Berdasarkan gap penelitian tersebut, penelitian ini bertujuan untuk menguji keandalan arsitektur *Hybrid CNN-Transformer* dalam mengklasifikasikan citra *chest X-Ray pneumonia*, serta menguji ketahanan model terhadap input luar domain melalui skema kelas kontrol.

B. METODE PENELITIAN

Metode Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode *deep learning* untuk mengklasifikasikan citra *X-Ray* paru-paru. Penelitian dilakukan di RS Labuang Baji Kota Makassar dengan data citra *X-*

Ray periode 2023 sampai 2025. Unit analisis penelitian adalah citra X-Ray paru-paru, sedangkan kelas keluaran meliputi *non pneumonia*, *pneumonia*, dan *random* atau citra bukan X-Ray paru-paru sebagai kelas kontrol.

Dataset penelitian terdiri atas 450 citra. Data tersebut dibagi ke dalam tiga kelas dengan jumlah seimbang, yaitu 150 citra *pneumonia*, 150 citra *non pneumonia*, dan 150 citra *random*. Proses *preprocessing* dilakukan melalui *resize* citra, normalisasi nilai piksel, serta peningkatan kualitas citra menggunakan CLAHE. Tahap augmentasi dilakukan untuk memperkaya variasi data dan mengurangi risiko *overfitting* pada model.

Model yang diterapkan meliputi CNN sebagai *baseline*, *Vision Transformer* (ViT), dan *Hybrid CNN-Transformer*. Pada model hybrid, CNN digunakan untuk menangkap fitur lokal citra, sedangkan *Transformer* digunakan untuk memahami hubungan global antarfitur. Fitur lokal dan fitur global kemudian digabungkan sebelum masuk ke lapisan klasifikasi akhir.

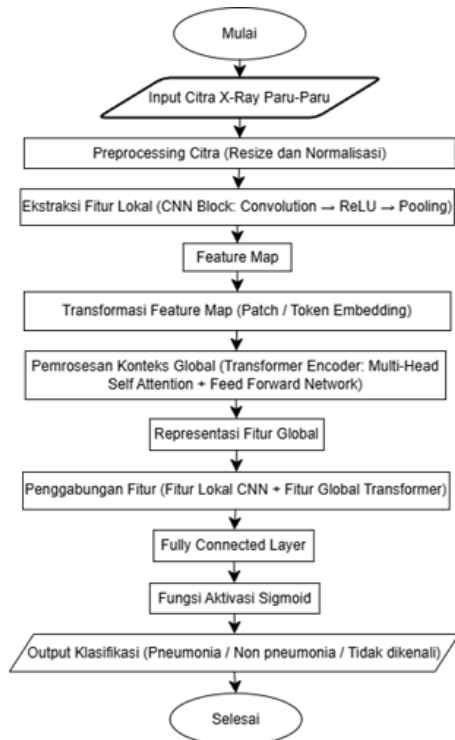
Table 1. Struktur Dataset Penelitian

Kelas	Jumlah Citra	Keterangan
Pneumonia	150 citra	Citra X-Ray paru-paru dengan indikasi pneumonia
Non pneumonia	150 citra	Citra X-Ray paru-paru tanpa indikasi pneumonia
Random	150 citra	Citra kontrol untuk menguji ketahanan sistem terhadap input di luar X-Ray paru

Total	450 citra	Keseluruhan data penelitian
-------	-----------	-----------------------------

Table 2. Konfigurasi Pelatihan Model

Parameter	CNN	ViT	Hybrid CNN-Transformer
Learning rate	8e-5	3e-5	5e-6 - 4e-4
Weight decay	1e-3	1e-3	5e-3
Batch size	32	32	32
Epoch	40	40	40
Optimizer	Adam W	Adam W	AdamW
Label smoothing	0,05	0,05	0,12



Gambar 1. Alur penelitian Hybrid CNN-Transformer

Evaluasi model dilakukan menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, *F1-score*, *AUC*, *ROC curve*, dan *Grad-CAM*. *Confusion matrix* digunakan untuk melihat jenis kesalahan klasifikasi. *Grad-CAM* digunakan untuk menilai area citra yang menjadi fokus model saat menghasilkan prediksi

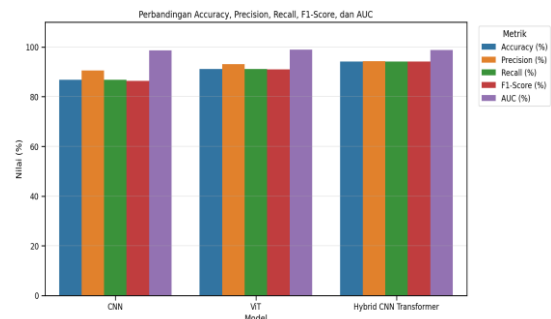
C. HASIL DAN PEMBAHASAN

Pengujian dilakukan untuk membandingkan kinerja CNN, ViT, dan *Hybrid CNN-Transformer*. Hasil pengujian menunjukkan bahwa setiap model mampu melakukan klasifikasi citra, tetapi model *hybrid* memberikan hasil paling baik pada sebagian besar metrik evaluasi.

Table 3. Confusion Matrix Model Hybrid CNN-Transformer

True Label	NP	R	P	Total
Non pneumonia	22	0	0	22
Random	3	20	0	23
Pneumonia	0	0	23	23
Total	25	20	23	68

Berdasarkan *confusion matrix*, model *Hybrid CNN-Transformer* mampu mengklasifikasikan seluruh citra *non pneumonia* dan *pneumonia* dengan benar pada data uji yang ditampilkan. Kesalahan terjadi pada tiga citra *random* yang diprediksi sebagai *non pneumonia*. Temuan ini menunjukkan bahwa model cukup kuat dalam membedakan citra paru normal dan *pneumonia*, tetapi masih memerlukan penguatan untuk mengenali input di luar domain X-Ray paru-paru



Gambar 2. Perbandingan kinerja CNN, ViT, dan Hybrid CNN-Transformer

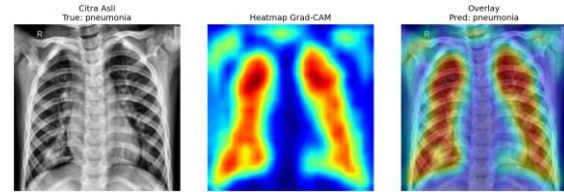
Model CNN memperoleh *accuracy* 86,76%, *precision* 88,03%, *recall* 86,76%, *F1-score* 86,58%, dan *AUC* 0,9858. Nilai ini menunjukkan bahwa CNN dapat mengenali pola citra X-Ray, tetapi masih memiliki keterbatasan dalam membaca konteks global. Keterbatasan tersebut dapat muncul ketika pola *pneumonia* terlihat halus atau tersebar pada beberapa bagian paru.

Model ViT memperoleh *accuracy* 94,12%, *precision* 95,02%, *recall* 94,12%, *F1-score* 94,12%, dan *AUC* 0,9948. Peningkatan ini menunjukkan bahwa mekanisme *self-attention* membantu model memahami hubungan antarbagian citra secara lebih luas. Namun, ViT tetap memiliki risiko kurang stabil pada dataset terbatas karena model ini cenderung membutuhkan jumlah data latih yang lebih besar.

Model Hybrid CNN-Transformer memperoleh hasil terbaik dengan *accuracy* 95,59%, *precision* 96,12%, *recall* 95,59%, *F1-score* 95,58%, dan *AUC* 0,9968. Nilai ini menunjukkan bahwa penggabungan fitur lokal dari CNN dan konteks global dari Transformer mampu menghasilkan representasi citra yang lebih informatif. Dibandingkan CNN, model hybrid meningkatkan *accuracy* sebesar 8,83%.

Table 4. Analisis Evaluasi Kinerja Model

Model	Plus	Mines	Penyebab Kesalahan Utama
CNN	Sederhana dan relatif cepat dilatih	Rentan salah saat pola citra samar	Fokus pada fitur lokal dan kurang menangkap konteks global
ViT	Akurasi tinggi	Berisiko overfitting pada dataset kecil	Masih sulit membaca pola infeksi yang sangat halus
Hybrid CNN-Transformer	Accuracy dan AUC tertinggi, performa paling stabil	Kompleksitas model lebih tinggi	Kesalahan tersisa terjadi pada citra di luar domain X-Ray paru



Gambar 3. Visualisasi Grad-CAM pada citra pneumonia

Visualisasi Grad-CAM menunjukkan bahwa model memberi perhatian pada area paru-paru yang relevan saat melakukan klasifikasi. Pada citra *pneumonia*, area yang mendapat bobot perhatian cenderung berada di bagian paru yang menunjukkan pola opasitas atau perubahan tekstur. Hasil ini mendukung interpretasi bahwa keputusan model tidak hanya berasal dari latar belakang citra, tetapi juga dari bagian anatomis yang berkaitan dengan indikasi pneumonia.

D. KESIMPULAN DAN SARAN

Model Hybrid CNN-Transformer berbasis MedViT berhasil diterapkan untuk klasifikasi penyakit *pneumonia* pada citra X-Ray paru-paru. Model ini menggabungkan kemampuan CNN dalam mengekstraksi fitur lokal dengan kemampuan Transformer dalam memahami konteks global citra.

Hasil pengujian menunjukkan bahwa model Hybrid CNN-Transformer memperoleh performa terbaik dengan *accuracy* 95,59%, *precision* 96,12%, *recall* 95,59%, *F1-score* 95,58%, dan *AUC* 0,9968. Model ini lebih baik dibandingkan CNN dan ViT pada metrik utama, terutama karena mampu menghasilkan representasi fitur yang lebih lengkap.

Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan lebih beragam, menambah validasi eksternal dari rumah sakit berbeda, serta

menguji metode *preprocessing* dan arsitektur model lain agar sistem klasifikasi dapat memiliki kemampuan generalisasi yang lebih kuat.

E. REFERENSI

- [1] Manzari, O. N., Ahmadabadi, H., Kashiani, H., Shokouhi, S. B., & Ayatollahi, A. (2023). MedViT: A robust vision transformer for generalized medical image classification. *Computers in Biology and Medicine*, 157, 1–15. <https://doi.org/10.1016/j.compbiomed.2023.106791>
- [2] Mathi, S., Deb, D., Chaudhuri, A. K., Joseph, L., & Biswas, S. (2025). A Review Of Convolutional Neural Networks For Medical Image Analysis: Trends And Future Directions. *Journal of Neonatal Surgery*, 14(7), 90–97. <https://doi.org/10.63682/jns.v14i7.5117>
- [3] Rahmadewi, P., Yanti, J., Harahap, K., & Indra, E. (2023). Algoritma Deep Learning untuk Pengklasifikasian Penyakit Radang Paru-Paru pada Citra Chest X-Ray dengan Convolutional Neural Network diderita pasien (Klasifikasi Penyakit Paru-Paru Berbasis Pengolahan Citra X Ray Menggunakan Convolutional Neural Network. *Jurnal Teknologi Informatika Dan Komputer MH. Thamrin*, 9(1), 14–23.
- [4] Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., Aoyama, R., Teraya, N., Bolatkan, A., Shinkai, N., Machino, H., Kobayashi, K., Asada, K., Komatsu, M., Kaneko, S., Sugiyama, M., & Hamamoto, R. (2024). Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. *Journal of Medical Systems*, 48(1), 1–22. <https://doi.org/10.1007/s10916-024-02105-8>
- [5] Taula'bi, W. K. S., Wahyudin, B., & Hariadi, A. (2024). Karakteristik Pneumonia Pada Balita Di RSUD Labuang Baji Makassar Tahun 2021 - 2022. *Bosowa Medical Journal*, 2(1), 29–33. <https://doi.org/10.56326/bmj.v2i1.5012>
- [6] Zahra, M., & Rosfadilla, P. (2025). Muthia Zahra 1 , Puspa Rosfadilla 2 Fakultas Kedokteran Universitas Malikussaleh 1 , Departemen Ilmu Pulmonologi dan Respirasi RSUD Cut Meutia Aceh Utara 2. *Jurnal Riset Ilmu Kesehatan Umum*, 3(2), 147–153.
- [7] Althubiti, S. A., Al-Ataby, A., & Ashour, W. M. (2024). A hybrid CNN-Vision Transformer approach for enhanced pneumonia detection in chest X-ray images. *Diagnostics*, 14(3), 285. <https://doi.org/10.3390/diagnostics14030285>
- [8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Hounsfield, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*, 1–21. <https://doi.org/10.48550/arXiv.2010.11929>
- [9] Hasan, M. M., Islam, M. R., & Kibria, M. G. (2023). Lightweight hybrid CNN-Transformer model for automatic pneumonia classification

- from chest radiograph images. *Informatics in Medicine Unlocked*, 41, 101320. <https://doi.org/10.1016/j.imu.2023.101320>
- [10] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 1122–1131. <https://doi.org/10.1016/j.cell.2018.02.010>
- [11] Rajaraman, S., Candemir, S., Kim, I., Thoma, G., & Antani, S. (2018). Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiograph. *Applied Sciences*, 8(8), 1397. <https://doi.org/10.3390/app8081397>
- [12] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *IEEE International Conference on Computer Vision*, 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- [13] Siddiqi, R. (2024). Robust chest X-ray classification for pneumonia detection: Leveraging hybrid deep learning architectures. *Biomedical Signal Processing and Control*, 88, 105612. <https://doi.org/10.1016/j.bspc.2023.105612>